The British Journal for the Philosophy of Science

VOLUME IV AUGUST, 1953 No. 14

OBITUARY

HANS REICHENBACH

PROFESSOR HANS REICHENBACH died suddenly on 9th April 1953, at his home in California. He was 62 years old.

Hans Reichenbach was born in Hamburg (Germany). His studies at various Universities were interrupted by war service; he obtained his Ph.D. degree at Erlangen with a thesis on *Relativitätstheorie und Erkenntnis a priori* (published in 1920). His first interest was, then, relativity theory, and on this subject he wrote many papers and two books, *Die Axiomatik der relativistischen Raum-Zeitlehre* (1924) and *Die Philosophie der Raum-Zeitlehre* (1928). These books established Professor Reichenbach's reputation and today have become classics; unfortunately, both works are now unavailable, and have not yet been translated into English. Popular lectures given during this time were collected in a volume entitled *Von Kopernicus zu Einstein* (about 1930), of which an American translation appeared in 1942. Another book, also written for the general public, called *Atom and Cosmos* was published about the same time in German and later, in French, English (1932), and American (1933) editions.

Meanwhile, Reichenbach had become *Privatdozent* in the faculty of science at the Technische Hochschule in Karlsruhe. From there, in 1928, he was called as *auserordentlicher Professor* to the University of Berlin, again in physics (which, however, was part of the philosophical faculty at that University). In Berlin, he attracted many students; and he was the leading member of the *Berliner Gesellschaft für wissenschaftliche Philosophie* which was, in some ways, akin to the Vienna circle. In 1930, he took over the journal *Erkenntnis* and, together with R. Carnap, made it into what was the first German periodical exclusively devoted to the philosophy of science.

In 1933, Reichenbach accepted a professorship at the University of Istanbul (Turkey). While there he published, in Holland, his *Wahrscheinlichkeitslehre* (1935) and, in the United States, the more popular *Experience and Prediction* (1938). The work on the frequency theory of probability and on induction marked the second phase of his development.

Late in 1938, after having served out his five-year contract in Turkey, Professor Reichenbach went to the University of California at Los Angeles. His teaching there was an unqualified success, so much so that his one-year course on logic and scientific method was made a prerequisite for all science students. In this, his third phase, Professor Reichenbach was occupied, apart from work in logic, mainly with quantum mechanics. His *Philosophic Foundations of Quantum Mechanics* (1944) was followed by *Elements of Symbolic Logic* (1947), and recently a popular book appeared under the title *The Rise of Scientific Philosophy* (1951). An English translation of the *Wahrscheinlichkeitslehre*, slightly modified and brought up-to-date was published as *The Theory of Probability* (1949).

In the early summer of 1952 Professor Reichenbach was invited to give lectures at the Institut Henri Poincaré (Paris) where, in 1937, he had also lectured on probability. His recent lectures were concerned with the concept of time and, I understand, a manuscript on this subject was practically finished at the time of his death.

His friends and pupils will feel the loss of a stimulating and original philosopher of science.

E. H. Hutten

ERRATA in May Issue, 1953

The date of Berkeley's birth on pp. i and 1 should read 1685, as on p. 84.

The portrait of Smibert has recently been cleaned, and now shows the date 1728, not 1725,

THE INTERPRETATION OF QUANTUM MECHANICS *

MAX BORN

THE following pages are a reply to Erwin Schrödinger's article, ' Are There Quantum Jumps? Parts I and II ', published in August and November 1952, in this *Journal*. A discussion on this subject was to be held in the meeting of the Philosophy of Science Group on 8th December 1952, and I was asked to open it. I accepted this honour rather reluctantly, for I find it awkward to display in public a disagreement on a fundamental question with one of my best and oldest friends. Yet I had several motives for accepting the challenge: The first is my conviction that no discrepancy of opinion on scientific questions can shake our friendship. The second, that other good and old friends of the same standing as Schrödinger, such as Niels Bohr, Heisenberg and Pauli, share my opinion. My third, and the most important reason for entering into this discussion of Schrödinger's publication is that by its undeniable literary merits, the width of its historical and philosophical horizon, and the ingenious presentation of the arguments, it may have a confusing effect on the mind of those who, without being physicists, are interested in the general ideas of physics.

The discussion on 8th December was rather frustrated by Schrödinger's absence, due to serious illness. I read my prepared introduction and answered questions. But this was, of course, not fair play to Schrödinger himself. Therefore I have to state my case in print. The following is a slightly enlarged version of my introduction to the discussion. As such, it covers not in the least all points made by Schrödinger, but only those which seemed to me suited for a debate amongst philosophers.

1 *Schrödinger's Case Restated*

The whole discrepancy is not so much an internal matter of physics, as one of its relation to philosophy and human knowledge in general. Any one of us theoretical physicists, including Schrödinger,

* Received 29.i.53

confronted with an actual problem would use the same, or at least equivalent mathematical methods, and if we should obtain concrete results our prediction and our prescription for the experimental verification would be practically the same. The difference of opinion appears only if a philosopher comes along and asks us : Now what do you really mean by your words, how can you speak about electrons to be sometimes particles, sometimes waves, and so on? Such questions about the real meanings of our words are just as important as the mathematical formalism. Schrödinger challenges the use of words in the current interpretation of the formalism ; he suggests a simple, puristic language and maintains that it can cope with the situation. We answer, that this purism is not only perfectly impracticable by its clumsiness, but also quite unjustifiable from the historical, psychological, epistemological, philosophical standpoint.

I suppose you have all read Schrödinger's paper. What he maintains can be condensed in a few sentences : The only reality in the physical world is waves. There are no particles and there are no energy quanta $h\nu$; they are an illusion due to a wrong interpretation of resonance phenomena of interfering waves. These waves are connected with integers in a way well known from the vibrations of strings and other musical instruments, and these integers have deluded the physicist into believing that they represent numbers of particles. But there is a special resonance law, characteristic of quantum mechanics, according to which the sum of the eigenfrequencies of two interacting systems remains constant. This has been interpreted by the physicists as the conservation law of energy applied to quanta or particles. But there are no such things. Any attempt to describe the physical phenomena in terms of particles without contradicting the well-established wave character of their propagation in space, leads to impossible, unacceptable conceptions, like the assumption of timeless quantum jumps of particles from one stationary state to another. Moreover, if you try to describe a gas composed of particles you are compelled to deprive them of their individuality ; if you write the symbol (A, B) to express that A is here at one place, B there at another, the two situations (A, B) and (B, A) are not only physically indistinguishable, but represent statistically only one case, not two, as common sense would demand. All these and many other difficulties disappear if you abandon the particle concept and use only the idea of waves.

2 *Are There Atoms?*

It is only a few years ago since Schrödinger published a paper under the title ' 2,500 Years of Quantum Mechanics ', in which he stressed the point that Planck's discovery of the quantum was the culmination of a continuous development starting with the Greek philosophers Leucippus and Democritus, the founders of the atomistic school. At that time he obviously thought the idea that matter is composed of atoms, ultimate indivisible particles, a great achievement. Now he rejects the same idea, because the execution of the programme leads to some grinding noise in our logical machinery.

It is this anti-atomistic attitude which appears to me the weakest, in fact quite indefensible, point in Schrödinger's arguments against the current interpretation of quantum mechanics. All other points are of a more technical nature, but this one is fundamental. Schrödinger opens both parts of his paper by a section entitled ' The Cultural Background ', in which he accuses the theoretical physicists of our time of having lost the feeling of historical continuity and over-estimating their own achievements as compared with those of their forerunners. He gives examples of such defaults which I do not wish to defend, but I think that he himself offers an example which is even worse.

The atomistic idea, since its revival through Daniel Bernoulli (1738) in the kinetic theory of gases and through Dalton (1808) in chemistry, has been so fertile and powerful that Schrödinger's attempt to overthrow it appears to me almost presumptuous, and in any case an obvious violation of historical continuity.

3 *Waves Instead of Atoms*

Such a violation would be justified if he could supply a better and more powerful substitute. That is exactly what he claims. He says that everything in physics and in chemistry as well can be described in terms of waves. The ordinary reader will certainly understand this as meaning : ordinary waves of some not specified substance in ordinary 3-dimensional space. Only in the last section of Part II (p. 241) he indicates that one has in general to do with waves in a multi-dimensional space, but ' To enlarge on this in general terms would have little value '. I think this is a very essential point which must be discussed. But before doing so I wish to say that I regard

Schrödinger's wave mechanics as one of the most admirable feats in the whole history of theoretical physics. I also know that his motive was his dislike of Bohr's theory of stationary states and quantum jumps, which he wished to replace by something more reasonable. I quite understand his triumph when he succeeded in interpreting those horrible stationary states as innocuous proper vibrations and the mysterious quantum numbers as the analogy to the numbers of musical overtones. He is in love with this idea.

I, of course, have no personal attachment to the waves. I have been involved, together with Heisenberg and Jordan, in the development of another method, matrix mechanics, in which stationary states and quantum jumps have a natural place. But I have no special preference for the matrix theory. As soon as Schrödinger's wave equation was published, I applied it to the theory of collisions ; this suggested to me the interpretation of the wave function as probability amplitude. I welcomed Schrödinger's elegant proof of the formal equivalence of wave mechanics and matrix mechanics. I do not plead in favour of matrix mechanics, or its generalisation due to Dirac, nor do I attack wave mechanics. I wish to refute the exaggerated claims of Schrödinger's paper from which the non-expert reader must get the impression that all phenomena can be described in terms of ordinary waves in ordinary space.

The physicist knows that this is not true. In the case of a 2-body problem (like the hydrogen atom) one can split the wave equation into two, one for the motion of the centre of mass, the other for the relative motion, both in 3-dimensional space. But already, in the case of the 3-body problem (for instance, the helium atom, one nucleus with two electrons) this is impossible ; one needs a 6-dimensional space for the relative motion. In the case of N particles one needs a 3(N-1)-dimensional space which only in singular cases is reducible to a smaller number of dimensions.

But this means that the claim of simplicity and of ' Anschaulichkeit ', the possibility of seeing the process in space, is illusory.[1] In fact a multi-dimensional wave function is nothing but a name for the abstract quantity ψ of the formalism, which by some of the modern

[1] In another article which has recently appeared (' Louis de Broglie, Physicien et Penseur ', ed. Albin Michel, Paris, 1952) Schrödinger remarks that the 3-dimensionality of the waves can be saved with the help of second quantisation. But the ' Anschaulichkeit ' is then also lost and the statistical character of the ψ-function is introduced on an even deeper and more abstract level.

theorists also goes under the more learned title of ' state vector in the Hilbert space '. Any attempt to describe phenomena, except the simplest ones, in terms of these multi-dimensional wave functions, means the formulation of the concise contents of mathematical formulæ in words of ordinary language. This would be not only extremely clumsy but practically impossible.

In fact, Schrödinger makes no attempt in this direction. All his examples are chosen in such a way that a 3-dimensional representation is possible. He restricts himself to cases which in the particle language correspond to independent (non-interacting) particles. Then he shows that these particles are not behaving as good, well bred particles, like a grain of sand, should behave.

4 *Why Atoms are Indispensable*

I think that in spite of these abnormities the concept of particle cannot be discarded.

As I said already, for the calculations of the theoretical physicist the whole question is almost irrelevant. But if he wants to connect his results with experimental facts, he has to describe them in terms of physical apparatus. These consist of bodies, not of waves. Thus at some point the wave description, even if it were possible, would have to be connected with ordinary bodies. The laws governing the motion of these tangible bodies are undoubtedly those of Newtonian mechanics. Thus the wave theory has necessarily to provide means to translate its results into the language of mechanics of ordinary bodies. If this is done systematically, the connecting link is matrix mechanics, or one of its generalisations. I cannot see how this transition from wave mechanics to ordinary mechanics of solid bodies can be possibly avoided.

Let us look at the matter the other way round, starting from ordinary bodies. These can be divided into parts, and sub-divided into still smaller parts. The Greek idea was that this procedure has an end somewhere, when parts become particles, atoms, which are indivisible.

Modern theory has modified this view to some degree, but I need not go into details which you all know. The parts of a substance obtained by division and subdivision are of the same physical nature until you approach the chemical atom. This is not indivisible, but its parts are of a different nature, particles of a more subtle quality,

nucleons and electrons. Then we discover that the smallest units, the chemical atoms and still more the nucleons and electrons have not only different qualities, but decidedly strange qualities, strange if you expect always to find the same as you are accustomed to. They behave differently from the powder particles into which you have first ground your material. They have no individuality, their position and velocity can be determined only with a restricted accuracy (according to Heisenberg's uncertainty relation) and so on. Shall we then say, well, there are no particles any more, we must regretfully abandon the use of this simple and attractive picture?

We can do it if we take a strictly positivistic standpoint: The only reality are the sense impressions. All the rest are 'constructs' of the mind. We are able to predict, with the help of the mathematical apparatus of quantum mechanics what the experimentalist will observe under definite experimental conditions, the current shown by a galvanometer, the track in a photographic plate. But it is meaningless to ask what there is behind the phenomena, waves or particles or what else. Many physicists have adopted this standpoint. I dislike it thoroughly, and so does Schrödinger. For he insists that there is something behind the phenomena, the sense impressions, namely waves moving in a still scantily explored medium. Recently an American physicist, Bohm, has taken the opposite standpoint; he claims that he can interpret the whole of quantum mechanics in terms of ordinary particles with the help of parameters describing unobservable 'concealed' processes.

5 *How to Modify the Atomistic Concept*

I think that neither of these extremist views can be maintained. The current interpretation of quantum theory which tries to reconcile both aspects of the phenomena, waves, and particles, seems to me on the right way. It is impossible to give here an account of the intricate logical balance. I wish only to illustrate the manner in which the particle concept is adapted to new conditions, by some examples from other fields, where a similar situation is found. It is of course no new situation that a concept in its original meaning turns out too narrow. But instead of abandoning it, science has applied another method, which is by far more fertile and satisfactory. Consider the example of the number concept. Number means originally what we now call integer, 1, 2, 3 . . . Kroneker has said that God has

made the integers, while the rest are human work. Indeed, if you define numbers as the means of counting things, even rational numbers like 2/3 or 4/5 are not numbers any more. The Greeks extended the concept of number to them by restricting the consideration to a finite set where a smallest unit (the greatest common denominator) can be found. But then they made the fundamental discovery that the diagonal of the square (of the side 1), which we write $\sqrt{2}$, is not a number in this sense; but great as their logical genius was, they did not make the next constructive step. They had not the pluck to generalise the number concept in such a way that $\sqrt{2}$ was included, but invented an ingenious yet rather clumsy geometrical method to deal with such cases. This was the stumbling block which retarded mathematics for about 2,000 years. Only in modern times the necessary generalisation of the idea of number was made so as to include these things such as $\sqrt{2}$, still called irrational. But then further generalisations followed, the introduction of algebraic, transcendental, complex numbers. You cannot count with the help of these. But they have other, more formal properties in common with the integers, and the latter are a special case. Similar generalisations of concepts are common in mathematics. But they appear also in physics. Sound was certainly defined as that which you can hear, light as that which you can see. But we speak now of inaudible sound (ultrasonics) and invisible light (infrared, ultraviolet). Even in ordinary life this process of extension of meaning is going on. Take the concept of democracy which originally meant the organisation of government in the Greek city states where the citizens assembled in the market place to discuss and decide their problems; today, it is used for the government of gigantic states by parliamentary representation. In Russia it even means something which we should regard as the opposite of democracy. Therefore we had better return to the safe ground of science.

I maintain that the use of the concept of particles has to be justified in the same way. It must satisfy two conditions: First it must share some (not in the least all) properties of the primitive idea of particle (to be part of matter in bulk, of which it can be regarded as composed), and secondly this primitive idea must be a special, or better, limiting case.

Now it is exactly in this sense that the particle concept is used in quantum mechanics. I cannot see any objection to it. Schrödinger's

examples seem to me of the kind which prohibited the Greeks from admitting the representation of the diagonal of the unit square as a number ; it differs from all possible ratios of integers, as can easily be seen. The effect of accepting Schrödinger's thesis would perhaps not be equally portentous, because he does not attack the formal theory, only its philosophical background. He would even allow the physicists and chemists to use the particle language with a proper 'as if'. Imagine a textbook of chemistry written according to this prescription. Water behaves as if it were composed of molecules H_2O, each of which again reacts as if composed of two H-atoms and one O-atom. But when we continue, each H-atom has properties as if it were composed of a nucleus and an electron, we transgress the permitted domain of 'as if', for here Schrödinger insists that there is no particle called electron but a charged wave around the nucleus which itself actually is also a wave of some kind. But when we then wish to deal with a photo-ionisation of this H-atom we have to fall back to his 'as if' to describe the discontinuous recording of a Geiger counter.

All our language, in life and in science, is growing through generalisations of concepts, which sometimes are first considered to be 'as ifs', but then are amalgamated and become legitimate words in their own right. For this end it is necessary to fix the rules of their employment in a reasonable manner. This process, in which Niels Bohr has played a leading part, is still going on, and, I think, with fair success. One can, of course, pick out points where some logical hardness or roughness appears, and that is what Schrödinger has done.

On the other hand, Schrödinger cannot avoid the use of the words particles or atoms. They appear in many of his examples ; otherwise his words would convey no meaning. For instance, when he speaks about quantum statistics of gases he has to discuss a wave equation in a multi-dimensional space. This equation has, of course, a simple meaning if considered from the particle standpoint ; it is the wave-mechanical translation of the law of conservation of kinetic energy for n particles. Now Schrödinger is compelled to disown this translation, the lovely child of his brain, for otherwise he would admit there are, in some sense, particles. He has to take the $3n$-dimensional wave equation as something given to him by inspiration and confirmed by experiments. This is a distortion of historical facts.

6 *Collisions*

Though I wish to avoid technical details I have to say a few words about the problem of collisions which Schrödinger discusses at several places (Sections 6 and 8). He finds the usual quantum-mechanical treatment faulty, he accuses the physicists of loose speech, he preaches to them that ' Science is not a soliloquy ' and prophecies that their work will be forgotten in 2,000 years' time, while that of Archimedes of Galileo, has survived similar periods. In a letter to me he maintains that ' almost all great successes of quantum mechanics consist of the satisfactory calculation of extended systems of eigenvalues (of the energy), each from a definite, more or less plausible assumption about the nature of the system in question (Hamilton operator), and have nothing at all to do with the statistical interpretation. On the other side there are the scattering experiments (calculation of differential cross-sections of interaction) and things like that. Only the Klein-Nishina formula is apparently quantitatively confirmed. (The latter represents the scattering of light, or photons, by an electron.) ' He further doubts that the statistical interpretation, which I have first suggested and which has been formulated in the most general way by von Neumann, is applicable to these cases at all.

To this I reply that in principle we know about the eigenvalues of the energy (Hamiltonian) of material systems only from experiments about emission, absorption, scattering of light or electrons. These processes are all due to the coupling of the system considered with a ' messenger ' field (the electromagnetic or photon field, or de Broglie's electron field) and it seems to me quite arbitrary to pick out the scattering as less reputable than the other two effects. Further, a look into the literature, for instance, the well-known book by Mott and Massey, or the important articles by Niels Bohr, on the penetration of particles through matter and innumerable other papers and books, shows that the number of more or less quantitative confirmations of the quantum-statistical scattering laws is very large, and that there are qualitative confirmations of a particularly convincing kind. Even in nuclear physics, where the knowledge of the interaction law (Hamiltonian) is doubtful and scanty, the principles of the statistical theory have been used with great success, of which the atomic bomb is one very impressive example.

Concerning Schrödinger's scepticism about the applicability of the general scheme for transitions (quantum jumps) to the case of

collisions I am unable to follow his reasoning. He describes the procedure as if a collision were a transition between two states of different energy. In fact the typical ' elastic ' collision is a transition between states of equal energy but different momentum vectors. My original method dealing with this case avoids any reference to time ; it considers the steady state of an incoming wave (representing a beam of ' messenger ' particles), transformed by its interaction with an atom into a spherical wave (representing the out-going, scattered particles). In this way of considering the process there is no initial and no final state, concepts which seem to Schrödinger ill-defined. They appear in Dirac's version of the collision theory which he developed in order to consider collisions as a special case of the general theory of transitions in time (formulated first in my papers on ' adiabatic invariants ' and in Dirac's simultaneous publications, and perfected by J. v. Neumann). But Dirac has shown that his method (involving time) is mathematically equivalent to the ' stationary ' method ; the conceptual difficulties which worry Schrödinger are therefore only a matter of careful formulation.

Another objection which he raises refers to the approximation method which I introduced in my early papers to solve the very complicated mathematical equations of scattering. This method gives reasonable, and often well-confirmed, results in the first approximation ; but higher approximations are difficult to obtain, and if they can be constructed there are cases where they lead to divergent integrals. However, there are other methods which use quite different expansions (for instance, in terms of spherical harmonics and Bessel functions) and lead to results which are mathematically sound and well confirmed by experiments.

I cannot see at all that these purely mathematical objections have anything to do with the question of ' particles-waves ', or ' quantum jumps '. For if we accept Schrödinger's standpoint that there are no particles, only waves, the scattering calculations would be exactly the same as before ; the only difference would be that we would speak about the intensity of the incoming and the outgoing wave (electromagnetic, electronic, protonic, etc., wave, as the case may be), and omit to interpret this intensity as the probability of the appearance of particles. The real problem raised by Schrödinger is, whether this probability interpretation is significant. His mathematical scruples have nothing to do with it. To decide this significant question, consider, for instance, Rutherford's experiments about the scattering

of α-rays by nuclei. Here, by a kind of lucky mathematical incidence, the classical calculation (using particles obeying the laws of Newtonian mechanics) and the wave-mechanical calculation (which can be performed rigorously in this case) give the same result. This result is confirmed by counting the α-particles in the incoming and in the outgoing beam (for different directions of scattering). The result is completely independent of the method of counting, whether by scintillations of a zinc-sulphide screen, or by different types of counters. How does Schrödinger account for this fact? As far as I see he has no ready explanation. He seems to think that it is not a discontinuity in the beam, which produces the countable events, but some feature of the counting instrument. But how then is it to be explained that the result is independent of the type of instrument, even to that degree, that sparks in the little crystals of the zinc-sulphide screen and gas tubes, connected with elaborate amplifier apparatus, count the same (average) number of events? Here Schrödinger's bias against the particle idea leads him to an almost mystical attitude; he hopes that the future will solve this riddle in a satisfactory way.

7 *Conclusion*

I have refrained from discussing the statistical interpretation of quantum mechanics in detail. This is not a simple matter, and demands not only the knowledge of a complicated mathematical formalism, but a certain philosophical attitude: the willingness to sacrifice traditional concepts and to accept new ones, like Bohr's principle of complementarity. I am far from saying that the present interpretation is perfect and final. I welcome Schrödinger's attack against the complacency of many physicists who are accepting the current interpretation because it works, without worrying about the soundness of the foundations. Yet I do not think that Schrödinger has made a positive contribution to the philosophical problems. It is very awkward for me to criticise the philosophy of a friend whom I deeply admire as a great scholar and deep thinker. Therefore I shall make use of a method of defence which Schrödinger himself is not too proud to use, namely the quotation of authorities who share my own opinion. I choose as my witness W. Pauli who is generally acknowledged to be the most critical, logically and mathematically exacting amongst the scholars who have contributed to quantum

mechanics. I translate a few lines from a letter (in German) which I have recently received :

> Against all retrograde efforts (Schrödinger, Bohr, etc., and, in a certain sense, also Einstein) I am certain that the statistical character of the ψ-function, and thus of the laws of nature—which you have, right from the beginning, strongly stressed in opposition to Schrödinger—will determine the style of the laws for at least some centuries. It is possible that later, for example in connection with the processes of life, something entirely new may be found, but to dream of a way back, back to the classical style of Newton-Maxwell (and it is nothing but dreams which those gentlemen indulge in), that seems to me hopeless, off the way, bad taste. And we could add ' it is not even a lovely dream '.

What Pauli means by the ' style ' of a conceptual structure you might prefer to call the philosophical attitude of a period, which determines the cultural background. It is here that we differ, and the auspices of an agreement are therefore frail.

Department of Mathematical Physics
The University, Drummond Street
Edinburgh 8

A VARIANT TO HILBERT'S THEORY OF THE FOUNDATIONS OF ARITHMETIC *

G. Kreisel

Summary

In Hilbert's theory of the foundations of any given branch of mathematics the main problem is to establish the consistency (of a suitable formalisation) of this branch. Since the (intuitionist) criticisms of classical logic, which Hilbert's theory was intended to meet, never even alluded to inconsistencies (in classical arithmetic), and since the investigations of Hilbert's school have always established much more than mere consistency, it is natural to formulate another general problem in the foundations of mathematics: to translate statements of theorems and proofs in the branch considered into those of some preferred system, where the translation must satisfy certain appropriate conditions (interpretation). The problem is relative to the choice of preferred system, as is Hilbert's consistency problem since he required the consistency to be established by particular methods (finitist ones).

A finitist interpretation of classical number theory, which has been published in full detail elsewhere, is here described by means of typical examples. Partial results on analysis (theory of arbitrary functions whose arguments and values are the non-negative integers) are here presented for the first time. One of these results is restricted to functions whose values are bounded; its interest derives from the fact that real numbers may be represented by such functions.

It is hoped that diverse general observations and comments, which would bore the specialist, may be of help to the general reader. The specialist may find some points of interest in the last two sections of the main text and in the notes following it.

1 *Introduction*

Some people develop preferences for certain methods in mathematics and dislikes of others; of course, different methods may be chosen by different people. The explanations which are offered may be atrocious: the methods which are disliked are variously 'doubted', called 'obscure', or, perhaps, claimed to be 'inappropriate' for certain

* Paper read to the Philosophy of Science Group on 9th February 1953

unspecified purposes. But generally there is a very good reason in the offing: we prefer the methods which come natural to us, and dislike those which don't. The facts just described suggest a number of interesting mathematical problems.

Above all, the chosen methods, and the others too, have to be so described that precise work can be done on them. Usually this is achieved by introducing formalisations of whole branches of mathematics: these make the discussion of our subject more *systematic* than does the pre-logistic use of allegedly typical examples. Once satisfactory[1] formatisations are available, we have to show that in some natural sense the chosen methods can be used to replace the others. The technique of replacing them is called 'interpretation'; more precisely, below I will state under what conditions such a technique may be called an interpretation. Lastly, when we have the chosen methods firmly in our minds, we may try to invent *rationalisations* for the preference; for instance, we may try to discover some specially interesting characteristics which single out these methods.[2]

It will be clear that I follow the general spirit of HB[3] very closely.

[1] 'Satisfactory' means that the properties of the informal branch which are under discussion, have really been represented by means of the formal system. For, trivially, such a branch—i.e. the (finite number of) arguments which have been put down on record or thought out up to any given time—may be embedded in diverse formal systems. And if the system proposed seems artificial, mathematicians will be (rightly) unmoved even by an inconsistency in the system; just as physicists are not bothered by the breakdown of a rule in an unintended application even if the latter was not excluded explicitly in advance.

[2] It is natural that we should use chosen methods in building a foundation, for no system of concepts would be considered a foundation for a branch of mathematics if they invited questions concerning themselves which are analogous to those they are intended to answer. This is true not only of the foundations of mathematics, but also of explanations in the fundamental sciences (Dirac). A foundation which consists of replacing the methods in a given branch by orthodox (or, more generally, preferred) ones is satisfactory from our present point of view since there is an end to this procedure, namely, when the latter ones have been reached. There are philosophers who insist on treating all mathematics as 'comparable' activities (the anthropologist's approach): then, of course, there can be no mathematical foundation for any branch of mathematics. (Nor, incidentally, does this philosophy provide a foundation in our sense either, because we should inevitably regard this philosophical activity as a subject matter for anthropology, too.) It seems to me quite reasonable to scorn foundations; but even if we do, the work described below need not be without interest, because it establishes connections between different methods of proof.

[3] Hilbert-Bernays, *Grundlagen der Mathematik*, Vol. I (1934), Vol. II (1939)

The main difference concerns only the starting point: while Hilbert wanted a consistency proof, I aim at an interpretation. But, I think, it is generally agreed (even HB, Vol. I, p. 19) that a consistency proof does not constitute the most direct[1] foundation, at any rate for arithmetic and analysis.

I shall confine myself to arithmetic and analysis, not only because they are important branches of mathematics, but also because they are familiar and of long standing, and so we are not tempted into 'making words mean what we like'. Geometry has well defined concepts too, but its foundations seem to me so well under control that I shall not delay over it.

An elegant formalisation of analysis is given in Supplement IV to HB, Vol. II (system H). Current number theory is formalised in Z_μ (HB, Vol. II, p. 293).

Now, there are well-known difficulties with *reading*[2] the formulae of these systems; e.g. a formula $\mathfrak{A}[\mu_x \mathfrak{A}(x)]$ would be read as: there exists an integer x for which $\mathfrak{A}(x)$ holds. Naïvely one would suppose that if $\mathfrak{A}[\mu_x \mathfrak{A}(x)]$ has been proved in our systems, there should be a term $O, O', O'', \ldots$ (an integer) for which $\mathfrak{A}(O^{(n)})$ can be proved too (in the system considered). This is not so. Actually, I think the exposition will be clearer if we do not work up paradoxes here, but take a calmer line. On the one hand we have reasonable descriptions of methods which are actually current; on the other I shall propose simpler[3] methods for arithmetic (and I shall give my reasons for

[1] There is a more specific reason for preferring the interpretation problem: the consistency of certain systems, for instance, pure logic (with and without a distinction of types), may be established by showing that the proved formulae of the system represent true propositions in a domain with a single element; and this is just what was never questioned! The demand for an interpretation, especially a complete one, cannot be fobbed off in such an uninformative way. It should be noted, however, that the consistency proofs in HB are by no means trivial, in fact, they establish much more than mere consistency. So, why not say so?

[2] Brouwer's criticism was specially emphatic. A formal system which limits methods of proof in accordance with Brouwer's ideas is due to Heyting. Though it restricts considerably the 'classical' use of disjunction (*or*) and negation (*not*), implications with undecided premisses are rampant; not only single ones, but structures like $\{[(A \to B) \to C] \to D\} \to [E \to (F \to G)]$: to read them one has to peel off brackets in stages, one cannot, so to speak, take in the whole structure in one go. The quantifier-free methods which I use below, are more restrictive: every closed formula, that is one without free variables, is decidable.

[3] They constitute a subclass of the methods which have been described as *finitist*. It has often been remarked that there can be no characterisation of the totality of

choosing them): so let us see whether we can interpret the former methods by these simpler ones. The only thing to be got out of the way before we can start is a precise formulation of this notion of interpretation.

2 *Interpretation*

Let (F) denote a formal system which represents chosen methods, and $(\mathfrak{F})$ one which represents the branch which is to be given an interpretation (foundation). Two distinct problems arise:

(i) *Elimination.* If A is a formula which belongs to both (F) and $(\mathfrak{F})$, we wish to establish that if A can be proved in $(\mathfrak{F})$ it can also be proved in (F). (Hilbert's consistency formulation, for arithmetic, amounts to this, when A is restricted to be a numerical formula, that is one without variables.)

(ii) *Interpretation.* If $\mathfrak{A}$ is a formula of $(\mathfrak{F})$, but not of (F) then a straightforward elimination (of the methods of $(\mathfrak{F})$) is of course impossible. Instead one may try to associate systematically with each such formula $\mathfrak{A}$ a formula A of (F) such that

(α) *from a proof* $\mathfrak{P}$ *of* $\mathfrak{A}$ *in* $(\mathfrak{F})$ *one can read off a proof* $P(\mathfrak{P})$ *of* A *in* (F);

finitist methods by means of a formal system. This is true; but for the interpretation of a particular (non-finitist) system or group of systems, one needs only a subclass of finitist methods. And there is no reason why these should not be formalised in a tidy system. (I believe there is no serious difference between our quantifier-free proofs and the class of finitist proofs indicated in HB, Vol. I, pp. 20–36; however, we emphasise different properties of finitist proofs: cf. para. 2 and para. 6 (iv) of the *Conclusion* below.) I do not consider that 'finitist' and 'non-finitist' are good words to describe the distinction between the methods used in Hilbert's theory of proof and the (classical) methods which are the subject matter of Hilbert's theory. For the difficulties which are usually stressed, namely, the ambiguities in the interpretations of logical connectives, could also occur with finite sets, e.g. if we considered a property of elements in finite sets which is not systematically decidable. (For example: a finite set of reals, which may even be decimals defined by primitive recursive functions, and the ordinary magnitude relations.) It is an accident that in ordinary practice most undecidable properties are introduced by taking a decidable relation $A(n, m)$ between integers n and m and making the property $(Em)A(n, m)$ out of it, which need not be systematically decidable. And, in fact, the arithmetic of addition and subtraction of integers is quite *transparent*, though we are here dealing with the infinite set of integers. (All this does not mean that decidability of the properties used is a *necessary* condition for a proof to be transparent. I merely wish to stress that the *particular* objections which prompted Hilbert's theory of proof could occur with finite sets.)

(β) *from a disproof* $\mathfrak{P}$ *of* $\mathfrak{A}$ *in* ($\mathfrak{F}$) *one can read off a disproof* $P_1(\mathfrak{P})$ *of* A *in* (F);

(β') If (F) is sufficiently restricted one may even be able *to read off a proof* $\mathfrak{P}(P)$ *of* $\mathfrak{A}$ *in* ($\mathfrak{F}$) *from any proof* P *of* A *in* (F);

(α) and (β') together (the conditions for a *complete* interpretation) may be replaced by:

(CI) $\mathfrak{A}$ *can be proved in* ($\mathfrak{F}$) *if and only if*[1] A *can be proved in* (F).

Actually, in arithmetic the most natural choice for (F) seems to be a system without quantifiers: then there is no interpretation which satisfies conditions (ii) (α) and (β') above.[2] Instead one associates with

[1] '$\mathfrak{A}$ can be proved in ($\mathfrak{F}$)' or 'A can be proved in (F)' are not systematically decidable propositions unless the systems ($\mathfrak{F}$) and (F) are decidable. So our condition (CI) contains an implication with a possibly undecided premis, which is just what we objected to above (p. 109, n. 2). The objection does not apply to conditions (α), (β), (β') because for any given $\mathfrak{P}$ or P one can decide whether $\mathfrak{P}$ (or P) is a proof of $\mathfrak{A}$ (or A) in ($\mathfrak{F}$) (or (F)) with the systems which we consider. But CI can easily be reworded:

(β'') *if* $\mathfrak{A}$ *is not proved in* ($\mathfrak{F}$) *by the sequence of formulae* $\mathfrak{P}(P)$ *then* P *is not a proof of* A *in* (F);

and (α) together with (β'') is a strengthened form of (CI). Incidentally, we have here a particular case of an abbreviation which is often used in stating results about formal systems which have the form:

if $A(x)$ *holds for all* x *then* $B(a)$ (i).

($A(n)$, $B(n)$ decidable for each numeral.) Since the proof of (i) always yields a computable function $g(a)$ which restricts the range of the quantifier (namely, if $A(x)$ holds for all $x \leqslant g(a)$ then $B(a)$), this function $g(a)$ is suppressed in the statement of the result. *N.B.*—It is essential that $A(n)$ should be systematically decidable for each numeral (0, 1, 2, . . .). After noting this fact (*Journal of Symbolic Logic*, 1951, **16**, 247) I used the abbreviation freely in my papers in the *Journal of Symbolic Logic*, **16**, and **17**, and in *Mathematische Zeitschrift*, 1952, **57**.

[2] *J. Symbol. Logic*, 1952, **17**, 57, appendix II. There is a simple alternative argument if it is further required that $\mathfrak{A}$ should be provable from the quantifier-free formula A; this condition is satisfied by our interpretation. For, suppose we relate to $(Ey)A_0(y)$ the formula $A_1(b)$ with the free variable b, where we suppose that the system ($\mathfrak{F}$) contains all primitive recursive $A_0(b)$. Then, under our condition, we should have: $(Ey)[A_0(y) \mathrm{v} \rightarrow A_1(y)]$ (for each primitive recursive $A_0(b)$); if the system in which $\mathfrak{A}$ *follows* from A, is ω-consistent then, for each $A_0(b)$ there must be a numeral n for which $A_0(n) \mathrm{v} \rightarrow A_1(n)$ holds. What is more, the integer n can be found by a general recursive procedure since it is defined by the expression $\mu_y[A_0(y) \mathrm{v} \rightarrow A_1(y)]$. But this means that the formula $(Ey)A_0(y)$ is decided by deciding $A_1(n)$, which is impossible, since then we should have a systematic decision method for the class of formulae $(Ey)A_0(y)$.

each $\mathfrak{A}$ of $(\mathfrak{F})$ a sequence A_n such that

(α) from a proof $\mathfrak{P}$ of $\mathfrak{A}$ in $(\mathfrak{F})$ we can read off a proof $P(\mathfrak{P})$ of $A_n(\mathfrak{P})$ in (F);

(β) from a proof P of (any) A_n we can read off a proof $\mathfrak{P}(P)$ of $\mathfrak{A}$ in $(\mathfrak{F})$.[1]

The use of a *sequence* A_n instead of a single A is, I think, not surprising: after all, $(Ey)A(y)$ would be interpreted by the light of nature as

$$A(0) \text{ or } A(1) \text{ or } \ldots$$

Perhaps the surprising thing is that even formulae $\mathfrak{A}$ with alternating quantifiers do not need more than sequences of quantifier-free formulae for their interpretation.

3 *Arithmetic*

Most of the arithmetic we do at school consists of the algebraic manipulation of identities and inequalities between integers; induction is used relatively rarely. In the theory of sequences and continuous functions, in other words, in elementary analysis, induction and the least number principle are used constantly. The former methods may be formalised in the so-called predicate calculus of first order with suitable identities and inequalities as axioms, the latter are formalised, I think satisfactorily, in the systems Z and Z_μ of HB.

Now, what methods are we to consider as specially transparent, what methods are we to use as foundations for arithmetic? I for one have been particularly attracted by quantifier-free methods. A quantifier-free system includes the usual axioms for addition and multiplication

$$a + 0 = a,\ a + b' = (a + b)',\ a \cdot 0 = 0,\ a \cdot b' = (a \cdot b) + a,$$

the elementary (propositional) calculus with free variables, definitions and proofs by induction, but not only ordinary induction

$$\frac{A(0),\ A(n) \rightarrow A(n')}{A(n)};$$

also so-called transfinite induction, which is based on well-orderings $a <\cdot\, b$ for the integers. ($a <\cdot\, b$ is called a well-ordering if each decreasing sequence a_n—with $a_{n+1} <\cdot\, a_n$—is finite: below, we take $0 <\cdot\, a \vee a = 0$, 0 is the first element in each ordering that we consider.) For any function $f(n)$ (in applications $f(n)$ is introduced by a recursion),

[1] *Math. Zeitschr.*, 1952, **57**, 1-12

we have the principle:

$$\frac{A(0) \,\&\, . f(n) <\cdot\, n \vee n = 0 \,.\, \&\, . A[f(n)] \rightarrow A(n)}{A(n)}$$

The idea is this: take any number n, and consider the sequence $(n), = a_1, f[f(n)], = a_2$, generally $a_{r+1} = f(a_r)$. Since the ordering is a well-ordering and $f(n) <\cdot\, n$ unless $n = 0$ we must have a number m such that $0 = a_m$ $(<\cdot\, a_{m-1} <\cdot \dots <\cdot\, a_1)$. Substituting repeatedly in the premise we get

$$\begin{gathered} A(0),\ A(a_m) \rightarrow A(a_{m-1}) \\ A(a_{m-1}) \rightarrow A(a_{m-2}) \\ \vdots \\ A(a_1) \rightarrow A(n), \end{gathered}$$

and hence we have a proof of $A(n)$. For the present the notion of an *arbitrary* sequence must be accepted quite naïvely. It will be discussed below in the section on analysis where it fits in naturally. At any rate it is interesting to observe that all objections to the classical use of quantifiers may be replaced in this way by objections to a proof of well ordering.

As a matter of fact it turns out that such quantifier-free systems are sufficient for an interpretation of (a formalisation of) classical arithmetic (Z_μ and extensions), even if the principle of transfinite induction is restricted to relatively simple orderings $a \underset{p}{<} b$.[1] But there is, I think, a more general justification for choosing these quantifier-free methods for a foundation of arithmetic (not for the theory of sets of points !). What is arithmetic 'about'? Surely, the integers 0, 1, . . . and numerical relations between them; for these form the starting point of arithmetic, ontogenetically and phylogenetically speaking. And what is a formula $A(n)$ about? Again, we should say $A(0)$, $A(1)$, . . . And how do we check these $A(0)$, $A(1)$, . . ., these numerical relations? By substitutions and computations. *Now*, given a quantifier-free proof of $A(n)$ we have a simple technique of reading off from the proof a numerical check of $A(0)$, $A(1)$, . . . : substitute for the variables, and replace the inductions by sets of implications as described above. (Of course, the crucial point is that a finite set of implications will carry one back to 0, and, perhaps, most people are apt to take this for granted merely because one can see how

[1] *Math. Annalen*, 1940, 117, 162–194

to *start* on these substitutions : but I think we shall see below that there is no reason to object to the proofs which establish that those orderings which are actually used are well-orderings.) In contrast to this, with classical methods the step from a proof of $A(n)$ to a computational check of $A(0^{(n)})$ requires the much more complicated reduction procedures of Gentzen[1] or Ackermann.[2] (We have here the usual ambiguity whether 'finitist' should refer to a concept or to its extension : if we speak of a *class* of finitist proofs, it is improper to call proofs, which are formalised in Z_μ, non-finitist ; perhaps, we should call them 'less finitist', since the reduction procedures provide a less direct passage from the general proof to the numerical example than is possible with quantifier-free proofs.)

What, then, do we want of an interpretation of classical arithmetic by these quantifier-free methods ?

(i) From a proof of the quantifier-free formula $A(n)$ in Z_μ we want to read off a quantifier-free proof. How we get it is described in Note I : it is quite easy. Perhaps an even more direct method might be obtained from the work of Schütte[3] and Lorenzen.[4]

(ii) Now consider a quantified formula, say $(x)(Ey)(z)A(x, y, z)$. If one is very naïve one looks for a function $f(x)$ (a systematic method of calculation) such that $A[x, f(x), z]$ is correct for each pair of integers x and z.

This is simply not possible, and if one looks at the standard proof of a formula like $(x)(Ey)(z)[A(x, y) \vee \rightharpoondown A(x, z)]$ one can't really expect it. Now, what do these proofs actually look like ? Let us consider as an example a proof of the convergence[5] of monotone bounded sequences. Consider a distance function $d(a, b)$, define the boundedness of a sequence by : $d(a_0, a_n) < C$, the monotone character by : $d(a_n, a_m) + d(a_m, a_p) = d(a_n, a_p)$ for $n \leqslant m \leqslant p$. We prove (the variables range over the integers) :

$$(n)(Em)(p)[p \geqslant m \rightarrow nd(a_m, a_p) < 1].$$

Of course, in general, one cannot estimate how m depends on n. Instead, the argument runs :

[1] *Math. Annalen*, 1936, **112**, 493-565
[2] *Ibid.*, 1940, **117**, 162-194
[3] *Ibid.*, 1951, **122**, 369-389
[4] *J. Symbol. Logic*, 1951, **16**, 81-106
[5] I prefer the proof of convergence to a proof of the existence of a limit : after all, monotone sequences of rationals converge without having a limit !

If the theorem were false, there would be an n_0 and a function $p_0(m)$ such that

$$p_0(m) \geqslant m, \text{ yet } n_0 d(a_m, a_{p_0(m)}) > 1 . \text{(I)}$$

In particular, putting $m = m_0 = 0$, $m_1 = p_0(0)$, . . . $m_{k+1} = p_0(m_k)$, we have for any $k_0 > Cn_0$: (1) is false for some $m \leqslant m_{k_0}$. m_{k_0} depends on the number Cn_0 and the function $p_0(c)$: it is a *functional* with these arguments.

In fact, the case can be generalised: given $(x)(Ey)(z)A(x, y, z)$, say $\mathfrak{A}$, a counter example would be a number a and a function $f(y)$ which satisfy $\rightarrow \mathfrak{A}$; that is, $\rightarrow A[a, y, f(y)]$ should hold for all y. But if $\mathfrak{A}$ can be proved by classical methods, so can $(Ey)A(a, y, f(y)]$, and there is[1] an integer $\mathfrak{y}$ which satisfies $A[a, \mathfrak{y}, f(\mathfrak{y})]$: this $\mathfrak{y}$ is expressed by a *functional* $\mathfrak{y}(f, a)$. And the value of this functional marks a break-down in the proposed counter example. We get an interpretation (the no-counter example-interpretation) satisfying our conditions above as follows. A sequence of functionals $\mathfrak{y}_0(f, a)$, $\mathfrak{y}_1(f, a)$ is obtained such that

(α) if $\mathfrak{A}$ is proved in Z_μ (or certain extensions), then we can find an n for which $A\{a, \mathfrak{y}_n(f, a), f[\mathfrak{y}_n(f, a)]\}$, call it A_n, can be proved without the use of quantifiers;

(β) $\mathfrak{A}$ is proved from A_n simply by substituting suitable terms (of Z_μ) in A_n: $\mu_x(y)(Ez) \rightarrow A(x, y, z)$, say $\mathfrak{y}$, for a, and $\mu_z \rightarrow A(\mathfrak{y}, b, z)$ for $f(b)$.

Actually the matter can be simplified further by the elimination of function variables. As we shall see below the functionals which we use are so simple[2] that, broadly speaking, $A\{a, \mathfrak{y}_n(f, a), f[\mathfrak{y}_n(f, a)]\}$ is 'true' for arbitrary functions $f(b)$ if it is 'true' for those functions $\nu_p(b)$ whose values differ from zero for a finite number of arguments only. (To avoid the vague 'true': A_n can be proved from

[1] In so-called externally consistent systems: if $A(b)$ is primitive recursive and $(Ex)A(x)$ has been proved there is an integer $n(0, 1, \ldots)$ for which $A(n)$ holds.

[2] For arithmetic with induction we require ordinal recursive functionals of finite order; these are obtained (i) by recursive definitions of the form

$$f_1(n) = g_0\{n, f_1[f_0(n)]\}$$

after $g_0(n, m)$, $f_0(n)$ have already been introduced and $f_0(n) \underset{p}{<} n \vee n = 0$, $f_1(n)$ is the function introduced; (ii) by the iteration functional

$$I_c[f(c), 0] = f(0)\,; \; I_c[f(c), n+1] = \max f(x) \text{ for } x \leqslant I_c[f(c), n].$$

For arithmetic without induction much simpler types of functionals are used: the so-called functionals of the predicate calculus, see e.g. *Math. Zeitschr.*, 1952, **57**, 3.

$(a)(p)A\{a, \mathfrak{y}_n(\nu_{\mathfrak{p}}, a), \nu_{\mathfrak{p}}[\mathfrak{y}_n(\nu_{\mathfrak{p}}, a)]\}$, call it A_n', in all systems which satisfy certain natural conditions : see Note III.)

The question of comparing different interpretations has often been raised (e.g. Ackermann, *Zentralblatt* Sept., 1952, called for a comparison between my trivial[1] interpretation and those which are described above). Now, several points strike me as relevant (it is perhaps not entirely accidental that in these comparisons my favourite interpretation comes out top on every count!). But here are the criteria for what they are worth :

(i) since an informal branch is better represented by classes of systems than by a single system, an interpretation for a particular system should remain one even if, say, verifiable quantifier-free formulae of the system are added to its axioms ;

(ii) an interpretation should guarantee that the system satisfies suitable conditions, e.g. a system for arithmetic should be externally consistent, and a system which is only weakly consistent should not have an interpretation ;

(iii) the interpretation should be complete ;

(iv) the system (F) should be quantifier-free.

Clearly, intuitionists would regard condition (iv) as artificial. But A. Heyting has pointed out that, nevertheless, our interpretation is of interest to intuitionists, because it sharpens Gödel's intuitionist interpretation of Z_μ: he gave a translation of formulae $\mathfrak{A}$ of Z_μ into formulae $\mathfrak{A}_1$, which can be proved intuitionistically if $\mathfrak{A}$ can be proved in Z_μ ; each A_n (of the sequence associated with $\mathfrak{A}$) is as strong as $\mathfrak{A}_1$, and sometimes stronger than $\mathfrak{A}_1$: 'strong' in the sense of intuitionist logic, which is the relevant standard of comparison here.

Many people like quantifiers, and find their elimination futile. But the fact remains that whatever methods may be rejected by other people these quantifier-free methods are not ; yet they are sufficient for an interpretation of those formal systems which are, so to speak, the worst of their kind (among those excluding bound function variables). So, at least the *effect* of our restriction is not too radical. There is, I think, a valid objection to the *wording* : for though the restriction makes good sense with arithmetic, it gives, it seems, no hint on what is to be done with other branches of mathematics.

[1] *J. Symbol. Logic*, 1951, 16, 248–249

4 *Analysis* (*Arbitrary Functions*)

In the theory of sets of points (arbitrary sets or functions) the elimination of quantifiers seems quite unnatural. But I think we can retain our leading idea in the interpretation of arithmetic, if we formulate it as follows: proofs in Z_μ (classical arithmetic) of the formula $\mathfrak{A}$ may be split up into (i) a *number theoretical core*, the quantifier-free proof of A_n, which depends on the particular relation $A(a, b, c)$ considered, and (ii) a 'verbal' reformulation, the step from A_n or A_n' to $\mathfrak{A}$, which is the same for all relations $A(a, b, c)$: this step is purely 'logical'. Our next problem is this: what is one to take as the mathematical core of a proof about arbitrary functions? (All the functions considered below are functions of the non-negative integers, and their values are non-negative integers too.)

To fix ideas let us consider propositions about arbitrary functions $f(n)$ which are made up as follows: we consider a quantifier-free formula of the first order predicate calculus to which are added (i) the function symbol f, (ii) symbols for arithmetic constants and primitive recursive functions. Such formulae are decidable when the individual variables are replaced by integers 0, 1, . . ., and if f is replaced by a function f_0 whose values can be computed for a suitable number of arguments. Note particularly that no method of evaluating f_0 need be given in advance, *f_0 need not be general recursive*, e.g. $f_0[f_0(2)] + f_0(0)$ is defined if the values of f_0 have been decided for the three arguments 0, 2, and $f_0(2)$. If some of the individual variables in such a proposition are bound by *restricted* quantifiers, e.g. $(y)[y \leqslant x \rightarrow f[f(x)] < f(y)]$ the formulae remain decidable in the sense above: e.g. in our example, if $x = 3$ we need the following values of f : $f(0), f(1), f(2), f(3)$ and $f[f(3)]$. Let $A(f, b)$ denote any such formula with restricted quantifiers and free variables b and $f(n)$. I shall discuss only two cases:

(i) $\underset{f}{\Lambda}(x)A(f, x)$ and $\underset{f}{V}(Ex)A(f, x)$,

(ii) $\underset{f}{\Lambda}(Ex)A(f, x)$ and $\underset{f}{V}(x)A(f, x)$.

(Read '$\underset{f}{\Lambda}$': for all functions f; '$\underset{f}{V}$': there is a function f.) We consider a *reduction* of these cases, which replaces quantification over *all* functions by quantification over a suitably chosen enumerable set of functions.

In case (i) $\Lambda(x)A(f, x)$ is *equivalent* to $(p)(x)A(\nu_p, x)$ where ν_p denotes the function whose value for the argument n is the highest

power of the nth prime which divides p; ν_p is an enumeration of the functions whose values differ from zero for a finite set of arguments only. The equivalence is seen as follows: suppose $A(f_0, x_0)$ is false; to decide $A(f_0, x_0)$ we need the values of f_0 for a suitable *finite* set of arguments only, hence there is a p_0 such that $A(\nu_{p_0}, x_0)$ is false, too. We can also state the result by reference to formal systems: the equivalence holds in any system in which the above argument can be formalised (formalisation presupposes of course the occurrence of function variables). The introduction of formal systems may seem very artificial; we shall not delay over it, but discuss it at the end of Note III.

An interesting application of this result uses Kleene's[1] representation of computable functionals $\mathfrak{y}(f, n)$

$$f\{\mu_y(x)\{x \leqslant y \rightarrow T[f(x), y, n]\}\}.$$

Then the formula $A\{n, \mathfrak{y}(f, n), f[\mathfrak{y}(f, n)]\}$, see p. 115, line 19, reads

$$(x)\{x \leqslant m \rightarrow T[f(x), m, n]\} \;\&\; (y)(Ez)\{y < m \rightarrow . \, z \leqslant y \;\&\; T[f(z), y, n]\} . \rightarrow A\{n, f(m), f[f(m)]\}$$

with the (free) variables n, m, f, and bound variables x, y, z, restricted not to exceed m. Thus our discussion of case (i) establishes the equivalence of A_n and A'_n, which was asserted without proof above.

In case (ii) such an equivalence no longer holds; thus for each general recursive f there may be an x satisfying $A(f, x)$ although $(x) \rightarrow A(\mathrm{t}, x)$ can be proved (in Z_μ) when a suitable term t is substituted for f. To see this, consider a formula

$$(x)(Ey)(z)[A_0(x, y) \vee \rightarrow A_0(x, z)]$$

which has no general recursive *Erfüllung*. Then for each general recursive f_0 one can find integers x_0 and z_0 to satisfy $\rightarrow A_0[x_0, f_0(x_0)]$ and $A_0(x_0, z_0)$ yet

$$(x)(z)\{A_0[x, \mathrm{t}(x)] \vee \rightarrow A_0(x, z)\}$$

can be proved in Z_μ if $\mathrm{t}(x)$ denotes $\mu_y(z)[A_0(x, y) \vee \rightarrow A_0(x, z)]$. There is an important special case where a reduction has been effected: if the functions $f(x)$ are required to be *bounded* $[f(x) < M]$, then there is a term $\mathrm{t}(x)$ of Z_μ of the form $\mu_y(z)B(x, y, z)$ with primitive recursive B, such that $\underset{f}{A}(Ex)A(f, x)$ is *equivalent* to $(Ex)A(\mathrm{t}, x)$. The proof uses the *Unendlichkeitslemma*; it is given in Note III. An

[1] *Proceedings of the International Congress of Mathematicians, 1950*, publ. 1952, 1, 679–685

application yields Brouwer's well-known theorem[1] stating that *computable functionals with bounded arguments are uniformly continuous.*

No analogous reduction is known in case (ii) for unbounded functions.

Now, it seems to me that such reductions of theorems about *all* functions to theorems about an *enumerable* set of functions provide a natural attack on the foundations of analysis; such reductions have also been found important in modern research on analysis in the form of *covering principles.* But in special cases it is more natural to argue with function variables, for instance, in establishing the well-ordering character of certain orderings $a <\cdot b$, i.e.

$$\underset{f}{A}(Ex)\{\rightharpoondown [f(x+1) <\cdot f(x)]\}.$$

Let $a <\cdot b$ hold if, and only if, (i) a is even and b is odd, (ii) a and b are of the same parity and $a < b$.

Again, let $a <o\, b$ be a well-ordering (of the integers), and define $a <\bullet b$ by the following conditions:

(i) $a = 2^{a_1} + 2^{a_2} + \ldots 2^{a_n} - 1$, $b = 2^{b_1} + 2^{b_2} + \ldots 2^{b_m} - 1$, $a_{r+1} <o\, A_r$, $b_{r+1} <o\, b_r$,

(ii) $a_i = b_i$ for $i < i_0$, $a_{i_0} <o\, b_{i_0}$; or $a_i = b_i$ for $i \leqslant n$ and $n < m$.

The traditional proofs showing that $a <\cdot b$ and $a <\bullet b$ are well-orderings seem perfectly transparent although they use function variables.

Consider the ordering $a <\cdot b$ first. If $f(0)$ is even $f(x+1) <\cdot f(x)$ cannot hold for all $x \leqslant \frac{1}{2}f(0)$; if $f(0)$ is odd and $f(x+1) <\cdot f(x)$ holds for all $x \leqslant \frac{1}{2}[f(0) - 1]$, $f\{\frac{1}{2}[f(0) + 1]\}$ is even, etc. Note that the function f is not assumed to be computable.

Next consider the ordering $a <\bullet b$, and define $I(a_1)$ to mean that there is an infinite descending sequence $\{a_i'\}$, $a = a_0' \cdot\!> a_1' \cdot\!> \ldots$ (i) If each a_i' exceeds $2^{a_2} - 1$, i.e. is of the form $2^{a_2} + \ldots - 1$, then $I(a_2)$ holds, too. (ii) If $2^{a_2} - 1$ exceeds a_k' where $a_k' = 2^{a_1} + 2^{a_2} + \ldots - 1$ then $I(\alpha_1)$ holds. Thus $I(a_1) \rightarrow \cdot I(a_2) \vee I(\alpha_1)$, i.e.

$$I(a_1) \rightarrow (Ex)[x <o\, a_1 \;\&\; I(x)].$$

But $\rightharpoondown I(0)$ holds trivially. Thus we have

$$\rightharpoondown I(0) \;\&\; .\, (x)[x <o\, a \rightarrow \rightharpoondown I(x)] \rightarrow \rightharpoondown I(a).$$

By appplying transfinite induction with respect to the ordering $a <o\, b$ to the (generally) *undecidable* predicate $\rightharpoondown I(a)$ we have

$$\underset{f}{A}(Ex)\{\rightharpoondown [f(x+1) <\cdot f(x)]\}.$$

[1] *Math. Annalen*, 1927, **97**, 60–75

Now, the orderings $a \underset{p}{<} b$ introduced in the section on arithmetic are obtained from $a <\cdot\, b$ by successive applications of the step from $a <o\, b$ to $a <\cdot\, b$. So I was justified, I think, in claiming that the well-ordering character of the orderings actually used is properly established.

In a rather weak sense there can be no reduction to a basic enumerable set of functions for the property of *well-ordering*, if (i) the basic set $f_n(c)$ is to be enumerable in the system considered, and (ii) $a < b$ is an arbitrary ordering of the system, see Note IV ; if $a < b$ is required to be decidable the question of finding a basic set of terms in Z_μ is open.

The reducibility problem may be put more formally thus : a sequence of functions $f_1(a), f_2(a), \ldots$ is called a *strong base* for a system if the formula $\underset{f}{V}\mathfrak{A}(f)$ can be proved if, and only if, some $\mathfrak{A}(f_n)$ can be proved (in the system considered) ; it is a *weak base* if $\underset{f}{V}\mathfrak{A}(f)$ cannot be proved when each $\rightarrow \mathfrak{A}(f_n)$ can be proved. (The numerals constitute a weak base for any ω-consistent system.) If a system contains Hilbert's selection symbol, the set of *all* terms is a base (and the set of terms $\epsilon_y(y <\cdot\, a)$ is a base for the property of well-ordering). But what we want for a base is a set of *one valued* functions.

5 *Conclusion*

I have described problems and solutions in the order which is natural to me. Let us see how they bear on some traditional questions in the foundations of arithmetic, such as : Why is the law of excluded middle (or some other principle of proof) permissible ? or : What is proof anyway ?

I referred, without discussion, to numerical arithmetic (the ontogenetic and phylogenetic starting point). Then quantifier-free arithmetic was introduced, and the reason for doing so was this : *a quantifier-free proof is a schema for a set of numerical calculations ;* in other words, there is a simple step from such a proof to a calculation of one of the particular cases which the proof is ' about '. We need only learn to substitute the symbols 0, 0′, 0″, . . . (or : 0, 1, 2, . . ., whichever notation is used) for letters of the alphabet. All other methods of arithmetic which we considered, were reduced to those above by means of an interpretation. It will be noticed that, as a result of formalisation and the arithmetisation of formal systems, our work on these methods of proof was closely analogous to work in quantifier-free arithmetic.

A demand for a *justification* of some principle of proof—when its application is restricted to *given* systems of proof—is now naturally formulated as follows: (i) to find a quantifier-free interpretation for formulae proved by means of the principle (in effect, an elimination of the principle), and, perhaps (ii) to find a looser (wrong) formulation of the principle which makes the desired elimination impossible.

In view of the eliminations which are available, we may, I think, claim that today we should have a philosophy of arithmetic if we had a philosophy of numerical arithmetic; actually, not only of arithmetic but also of the various branches of mathematics which have been arithmetised satisfactorily. (A philosophy should explain what makes arithmetic *transparent* and *certain*.)

But what are the difficulties in a philosophy of numerical arithmetic? I think they are such that we do not even have a glimmering of a (non-trivial) philosophy of numerical arithmetic. Let me explain what I mean.

Of course, there are a hundred and one facts about numerical arithmetic which strike us as philosophically interesting: (i) we learn it at an early age; (ii) the notation of arithmetic may be standardised without producing serious limitations (strokes only or decimal notation only need be used); (iii) any problem in numerical arithmetic may be decided systematically; (iv) its proofs may be regarded as schemata for experiments (*Gedankenexperimente*): a proof is a set of instructions for attaching and detaching (real or imagined) strokes where the whole process can be visualised in its entirety (*ein übersehbarer Prozess*), as described by Bernays in HB, vol. I. But what I want from a philosophy of numerical arithmetic is an emphasis on those facts (about numerical arithmetic) which lead to a generalisation: for instance, it should emphasise those facts which would suggest a notion of 'clear proof' in the theory of sets of points. And none of the facts mentioned above does it.

They are so special that, if we made them defining conditions, many methods could not be called clear though they strike us as transparent: this really happened with transfinite induction (the culprit: the length of decreasing sequences of integers with a given first term may be unbounded when the ordering is transfinite, but it is always bounded in the natural ordering of integers); and thus the principle conflicts with (iv)! On the other hand, proofs of *every* formal system satisfy (ii)-(iv), at any rate in the sense that with the usual systems one can check systematically whether a given proof has

been constructed in accordance with the rules. Yet we should not call a branch of mathematics clear merely because it has been formalised!

(More precisely; such a branch would indeed be clear enough to permit us to formulate the consistency problem or the completeness problem; but it would not necessarily be clear enough for us to argue freely with it: and this is our main concern at the moment.)

Whatever his general principles, a moderately sane person will go on thinking about particular problems by the light of nature. Whatever I may think about analysis generally, I feel that the reductions described in the main text are in the right direction. Perhaps it is not reasonable to try to force at present a decision concerning analysis: after all, when people began working systematically on arithmetic, its mathematical development had already attained a fairly rigid shape, and one application had come to be regarded as fundamental, namely counting. With analysis one has no really strong feelings on many open questions: if the continuum hypothesis should turn out to be independent of the usual axioms of set theory, should we not go on investigating the consequences both of the hypothesis and of its negation? If Fermat's conjecture were found to be independent of one of the usual sets of axioms of number theory, we should unhesitatingly regard it as proved since for every quadruple of positive integers n, m, p, q, $(1, 2, 3, \ldots)$ we should have $n^{q+2} + m^{q+2} \neq p^{q+2}$. Only a large number of decisive results in analysis and, perhaps, the accident of finding a homogeneous field of application for large portions of analysis may obliterate the difference between arithmetic and analysis which is indicated in the present paragraph.

After all these heart-searchings, what facts should I make responsible for the special clarity of numerical arithmetic? I think, first, the possibility of enumerating in a very natural fashion *all* the problems of numerical arithmetic; and, second, the systematic decidability of these problems. I say this, fully realising that at least the second fact has no analogue in analysis: in other words, fully realising that my answer does not even give a hint for a philosophy of mathematics.

Note I

$\mathfrak{F}$ contains Z_μ.

Let *Prov* (a, b), $e(m)$ be primitive recursive expressions which satisfy the following conditions:

(i) *Prov* (a, b) holds if, and only if, a is the number of a proof in $(\mathfrak{F})$ of the expression with number b, a Gödel numbering of $(\mathfrak{F})$ being assumed;

(ii) $e(m)$ is the number of the negation of the expression with number m; $\alpha(n)$ is the number of $A(0^{(n)})$;

(iii) for any primitive recursive $A(b)$, there is a primitive recursive function $\pi(n)$ such that

$$\neg A(n) \rightarrow Prov\, \{\pi(n),\ e[\alpha(n)]\}$$

can be proved in recursive arithmetic with ordinary induction.

Note.—Condition (iii) is necessary since there are, of course, many definitions of primitive recursive type satisfying (i) and (ii): e.g. if $Prov\,(a, b)$ is one so is $Prov\,(a, b)\ \&\ U(a)\ \&\ U(b)$ where $(x)U(x)$ is irrefutable in Z_μ; (iii) will hold only for a suitable choice of definition. One such choice is that of [HB], where, by HB, Vol. II, pp. 312-324,

$$\neg A(n) \rightarrow (Ey)\ Prov\, \{y,\ e[\alpha(n)]\}$$

with the free variable n can be proved in Z_μ. But by going through the argument, one reads off a function $\pi(n)$, the number of an *evaluation* of $\neg A(0^{(n)})$, such that

$$\neg A(n) \rightarrow Prov\, \{\pi(n),\ e[\alpha(n)]\}.$$

Observe that $\pi(n)$ is the number of a sequence of formulae without quantifiers.

We take as our formulation (*cf.* the note above) of the consistency of $\mathfrak{F}$ the formula

$$\neg Prov\,(m,\ n) \vee \neg Prov\,[p,\ e(n)] \qquad (1)$$

Given a proof $\mathfrak{P}$ in $(\mathfrak{F})$ of $A(b)$, a proof in $(\mathfrak{F})$ of $A(0^{(n)})$ consists of $\mathfrak{P}$ with the formula $A(0^{(n)})$ attached; suppose this has the number $\lambda(n)$. Then

$$Prov\,[\lambda(n),\ \alpha(n)],$$

and hence from (1)

$$\neg Prov\, \{\pi(n),\ e[\alpha(n)]\}.$$

By (iii) $\qquad A(n).$

Note in passing that Ackermann[1] uses only transfinite induction *up* to ω_p ($\omega_1 = \omega,\ \omega_{n+1} = \omega^{\omega_n}$) to establish

$$P(m)\ \&\ P(p) \,.\rightarrow.\, \neg Prov\,(m,\ n) \vee \neg Prov\,[p,\ e(n)],$$

where $P(m)$ is a (suitable) arithmetisation of the proposition: the proof with number m contains $< P$ quantifiers. Thus, given a proof in $\mathfrak{F}$ of $A(b)$ with $< P$ quantifiers, both $\lambda(n)$ and $\pi(n)$ are numbers of sequences of formulae with $< P$ quantifiers, and hence $A(n)$ can be proved by recursive arithmetic with transfinite induction up to ω_p only.

Note II

Schütte, in a paper of exemplary lucidity,[2] produced a much more far-reaching elimination: following Gentzen he introduces a system where

[1] *Math. Annalen*, 1940, **117**, 162-194 [2] *Ibid.*, 1950, **122**, 47-65

the formulae of a proof get successively more complicated, in particular, the number of quantifiers and function symbols is not decreased by the application of any inference rule of the system. Then he shows how, given a proof of $\mathfrak{A}$ in the predicate calculus—or in number theory without induction, we get a proof of $\mathfrak{A}$ in his system.

To get an analogous result for number theory with induction both Schütte [1] and Lorenzen [2] introduced 'systems' with a rather strange rule: if $A(0)$, $A(0')$, . . . $A(0^{(n)})$ . . . for each numeral $0^{(n)}$ can be proved in the system, then $A(b)$ with the free variable b is also a proved formula of the system. I had some difficulty with their formulation, until I found a restatement; I may mention it since other people may find it helpful:

Suppose we have a system of number theory consisting of verifiable axioms, the inference rules of the predicate calculus, and a rule of the form: if $\mathfrak{B}[\mathfrak{A}]$ then $\mathfrak{A}(b)$ with the free variable b; it is supposed that if $\mathfrak{B}[\mathfrak{A}]$ has been proved in the system by m applications of this last rule, so can $\mathfrak{A}(0^{(n)})$. Then the system is consistent. Examples of $\mathfrak{B}[\mathfrak{A}]$ are:

(i) $\mathfrak{A}(0)$ & $(x)[\mathfrak{A}(x) \rightarrow \mathfrak{A}(x')]$, ordinary induction, here $\mathfrak{A}(0^{(n)})$ can be proved from $\mathfrak{B}[\mathfrak{A}]$ by n substitutions and *modus ponens*, i.e. without the use of induction.

(ii) $\mathfrak{A}(0)$ & $(x)\{(y)[y <\cdot\, x \rightarrow \mathfrak{A}(y)] \rightarrow \mathfrak{A}(x)\}$, transfinite induction with respect to $<\cdot$; if $<\cdot$ is 'seen' to be a well-ordering, each integer $0^{(n)}$ is accessible, and $\mathfrak{A}(0^{(n)})$ is proved by a *transfinite* 'sequence' of substitutions in $\mathfrak{B}[\mathfrak{A}]$. Here the consistency proof for the system is relative to the methods which enable one to prove that $<\cdot$ is a well ordering.

(iii) $\mathfrak{A}(0)$ & $(x)[f(x) <\cdot\, x \vee x = 0]$ & $(x)\{\mathfrak{A}[f(x)] \rightarrow \mathfrak{A}(x)\}$, modified transfinite induction with respect to $<\cdot$; here only a finite number of substitutions in $\mathfrak{B}[\mathfrak{A}]$ is used if $<\cdot$ is a well-ordering.

This reformulation destroys the analogy with the predicate calculus stressed in [3]: the function which represents the transfinite sequence of (ii) need not occur in $\mathfrak{A}(b)$ itself, i.e. a function symbol that is introduced in the proof drops out later on, and the same applies to $f(x)$ in (iii). I don't think it is sensible to insist on the analogy, because it is bound to be artificial; it conceals one of the most interesting facts of quantifier-free recursive number theory: there are formulae whose proof requires auxiliary functions.

More precisely: in primitive recursive number theory there are formulae $A(n)$ which cannot be proved in this number theory without introducing auxiliary functions, but can be proved if suitable auxiliary functions are introduced—by defining equations of primitive recursive type.

[1] *Math. Annalen*, pp. 369-389

[2] *J. Symbol. Logic*, 1951, **16**, 81-106

[3] *Math. Annalen*, 1950, **122**, 47-65

I should like to mention two examples.

(i) In the reduction procedures[1] the main point is this: proofs of Z are (well-) ordered, in the first place, according to the number of inductions used, in the second place, according to the number of quantifiers (roughly speaking); it is then shown that if a formula A without variables can be proved by a proof $\mathfrak{P}$ with variables, it can also be proved by a proof which precedes $\mathfrak{P}$ in the ordering. Further, if A is false it cannot be proved by a proof without variables. This argument shows that if A is false it cannot be proved in Z. But a formulation of the argument which uses the principle of infinite induction, does not make the ordering of proofs explicit though every mathematician would agree that this ordering was the crux of the problem.

(ii) An immediate consequence of Ackermann's work[2] is this: if $(x)(Ey)A(x, y)$, $A(a, b)$ primitive recursive, can be proved in Z and $\mu_y A(n, y)$ is ordinal recursive of order m (but not of lower order) then the proof contains at least m quantifiers, and there are such A where $\mu_y A(n, y)$ is of order greater than 2. Yet the principle of infinite induction masks this interesting distinction.

Note III

We need a few definitions:

$A(f, b)$ denotes a formula of the pure predicate calculus, to which are added the function symbol f, symbols for primitive recursive (or for computable) functions; f is the only function variable of $A(f, b)$; those variables in $A(f, b)$ which are bound are restricted to lie in a limited range. A simple, but useful consequence of our restriction is this: if the values of f are bounded, say $f(x) < M$, one can write down a primitive recursive (computable) function $\lambda(n)$ such that $A(f, x)$ is decided for all $x \leqslant n$ provided the values of $f(x)$ for $x \leqslant \lambda(n)$ have been decided. To see this use induction with respect to the number of symbols occurring in the terms contained in $A(f, b)$. (Here and below all individual variables range over the non-negative integers.) The crux is that $\lambda(n)$ is independent of f.

The function $s(m, i, n)$ is introduced (uniquely) by the conditions:

$$m < M^{\lambda(n)},\ m = \sum_{i=1}^{\lambda(n)} s(m, i;\ n) M^{\lambda(n)-i},\ s(m, i;\ n) < M,$$
$$i \geqslant \lambda(n) \rightarrow s(m,\ i;\ n) = 0.$$

(Recall the familiar representation of integers by the M-ary scale.) m will be called the number of the sequence $s(m, 1;\ n), \ldots, s[m, \lambda(n);\ n] \cdot A_n(m, b)$ denotes the expression obtained by substituting the function $s[m, i;\ n)$ for $f(i)$ in $A(f, b)$.

[1] *Math. Annalen*, 1940, **117**, 162-194; 1936, **112**, 493-565

[2] *Ibid.*, 1940, **117**, 162-194

The Unendlichkeitslemma

There is a term $\mathfrak{t}(i)$ *of the form* $\mu_y(z)B(i, y, z)$, *where* $B(a, b, c)$ *is primitive recursive, such that*

$$(n)(Em)(x)[m < M^{\lambda(n)} \,\&\, . \, x \leqslant n \rightarrow A_n(m, x)] \rightarrow (x)A(\mathfrak{t}, x).$$

(The proof can be formalised in Z_μ, see e.g. HB, Vol. II, pp. 240–243, but in weaker systems too.)

$J(n, m, n', m')$ means : $n' \geqslant n$ and $(x)[x \leqslant n \rightarrow s(m, x\,; n) = s(m', x\,; n')]$. i.e. m represents the first n terms in the sequence with number m' and the latter sequence contains no less than n terms. The terms in this sequence are, of course, less than M. Now, if (i) $x \leqslant n' \rightarrow A_{n'}(m', x)$ and (ii) $J(n, m, n', m')$ we have

$$x \leqslant n \rightarrow A_n(m, x).$$

One then defines $\mathfrak{t}(n)$ by the expression

$$\mu_m(n')(Em')\{[J(n, m, n', m') \,\&\, m' \leqslant M^{\lambda(n')} \,\&\, . \, x \leqslant n' \rightarrow A_{n'}(m', x)] \vee n' < n\}.$$

The implication above follows by standard methods.

Applications

(i) $(x)[f(x) < M] \rightarrow . \underset{f}{V}(x)A(f, x) \longleftrightarrow (x)A(\mathfrak{t}, x)$ (with *analogous* $\mathfrak{t}$).

The proof of the equivalence is possible in any system in which (i) the functions $s(m, i\,;\, n)$ and $\mathfrak{t}(i)$ occur, (ii) substitution of $s(m, i\,;\, n)$ for a free function variable $f(i)$ is permitted. Thus in the case of *bounded* functions we have a reduction of propositions of the form

$$\underset{f}{A}(Ex)A(f, x) \text{ or } \underset{f}{V}(x)A(f, x)$$

to one about terms in Z_μ only.

In particular, the reduction applies to normal forms of functionals whose arguments are bounded functions : here one may either use Kleene's normal form[1] or, the much older equivalent form due to Brouwer.[2]

(ii) $(x)[f(x) < M] \rightarrow . \underset{f}{A}(Ex)A(f, x) \longleftrightarrow$

$$(En)(m)(Ex)[m < M^{\lambda(n)} \rightarrow . \, x \leqslant n \,\&\, A_n(m, x)].$$

The equivalence is seen as follows : by (i)

$$\underset{f}{A}(Ex)A(f, x) \rightarrow (Ex)A(\mathfrak{t}, x).$$

By the lemma

$$(Ex)A(\mathfrak{t}, x) \rightarrow (En)(m)(Ex)[m < M^{\lambda(n)} \rightarrow . \, x \leqslant n \,\&\, A_n(m, x)].$$

Thus one half of the equivalence is proved. Note particularly that in the conclusion the variables m and x are restricted to a bounded range : thus the conclusion is equivalent to a formula $(En)A^*(n)$ where $A^*(n)$ is *primitive recursive*.

[1] *Proc. Internat. Congr. Math.*, 1950, **1**, 679-685

[2] *Math. Annalen*, 1927, **97**, 60–75

The other half is seen as follows: if $A^*(n)$ holds and f is an arbitrary function whose values are $< M$, consider its first $\lambda(n)$ terms; the number of the sequence $f(1), \ldots, f[\lambda(n)]$ is $f^*, = \sum_{i=1}^{\lambda(n)} f(i)\, M^{\lambda(n)-i}, < M^{\lambda(n)}$. But $A^*(n)$ states that there is an $x^* \leqslant \lambda(n)$ which satisfies $A_n(f^*, x^*)$. By the first paragraph of this note, $A(f, x^*)$ must hold.

(The proof of (ii) applies in systems satisfying the conditions stated in (i).)

(ii) leads to a statement of Brouwer's theorem on the uniform continuity of computable functionals with bounded arguments:

Informal statement. Let a computable functional be represented by $f[\mu_x A(f, x)]$, where $(Ex)A(f, x)$ holds for arbitrary f with values $< M$; then there is an n such that $\mu_x A(f, x) \leqslant n$ if $(x)[f(x) < M]$. Thus the value of the functional is determined if the values of the function f are known for arguments $\leqslant n$.

Formal statement. Suppose $(\mathfrak{F})$ is any system which satisfies the following conditions:

(i) $(Ex)A(\mathfrak{t}, x)$ can be proved in $(\mathfrak{F})$;

(ii) the Unendlichkeitslemma can be proved in $(\mathfrak{F})$;

(iii) $(\mathfrak{F})$ is weakly ω-consistent, i.e. $(Ex)A^*(x)$, for primitive recursive A^*, can be proved in $(\mathfrak{F})$ if, and only if, there is an integer $0, 1, \ldots$ which satisfies A^*b.

Then we can find an integer $0^{(n)}$, by checking $A^*(0), A^*(1), \ldots$, i.e. by a systematic quasi-recursive procedure, such that to each of the $M^{\lambda(n)}$ possible sequences $s(m, i;\ n)$ there is an integer $x_m \leqslant \lambda(n)$ which satisfies $A_n(m, x_m)$.

What are the advantages of the formal statement over the informal one?

First, it is stronger than the informal one: (since such a comparison makes sense only if one is uncritical of the informal notion of a function variable we shall use this notion freely in the present paragraph). In general, a system $(\mathfrak{F})$, all of whose theorems represent intuitively true propositions, may remain weakly ω-consistent when an axiom $\mathfrak{A}$ is added, yet the formula $\mathfrak{A}$ may be intuitively false. In our case, we *assume* only that the addition of $(Ex)A(\mathfrak{t}, x)$ leaves such a system weakly ω-consistent, and *deduce* that $\mathfrak{t}(n) = 0$ and that no intuitively false formulae can be proved in the enlarged system.

Secondly, I, for one, cannot formulate (to my own satisfaction) conditions on proofs about arbitrary functions which ensure that these proofs are 'clear'. On the other hand, a proof of weak ω-consistency of a formalised system $(\mathfrak{F})$ establishes (after a suitable Gödel numbering has been introduced) a theorem of arithmetic of the form $(n)(Em)P(n, m)$ or $P[n, g(n)]$ where P is primitive recursive and g is computable. This

proposition contains no reference to arbitrary functions, and I know what I am after: rightly or wrongly, I want a quantifier-free proof of $P[n, g(n)]$.

Note. The reduction is interesting because real numbers may be represented by functions whose values are bounded.

Note further that the result is 'best possible': we exhibit a formula $\underset{f}{V}[(x)\,(Ey)A(f, x, y)\ \&\ f(x) = 0 \vee f(x) = 1]$ which can be proved in the usual systems of analysis, but for each term $\mathfrak{e}$ of Z we have $\rightarrow (x)\,(Ey)A(\mathfrak{e}, x, y)$ in Z (since

$$\{\epsilon_f\,(x)\,(Ey)A(f, x, y)\}\,(n) = 0$$

is a truth-definition for prenex formulae of Z).

We refer to HB, Vol. II, pp. 331-338, for definitions of the following primitive recursive predicates and functions (where n denotes the number of the formula N of Z):

$$S(n),\ S_0(n),\ M_0(n),\ U(n),\ E(n)$$

mean (i) N is a prenex formula (of Z) without free variables, (ii) N is a numerical formula, (iii) N is a true numerical formula, (iv) the first quantifier of N is universal, (v) the first quantifier of N is existential: $p(n, m)$ is the number of the formula obtained by deleting the first quantifier of N and replacing the corresponding variable by the numeral $O^{(m)}$.

Then we take for $A(f, x, y)$ the formula

$$\begin{aligned} &S_0(x) \rightarrow .\ f(x) = 0 \longleftrightarrow M_0(x) : \& : S(x_1)\ \&\ U(x_1).\ \rightarrow . \\ &\quad f(x_1) = 0 \longleftrightarrow (x_2)\{f[p(x_1, x_2)] = 0\} : \& : S(x_1)\ \&\ E(x_1).\ \rightarrow . \\ &\quad f(x_1) = 0 \longleftrightarrow (Ey)\{f\,[p(x, y)] = 0\}, \end{aligned}$$

where, in the notation of HB, Vol. II, p. 235, footnote 1, x_1 denotes $\eta_1(x)$, and x_2 denotes $\eta_2(x)$.

Note IV

Let $f_n(m)$ be a recursively enumerable sequence of functions. Define

$$g(1) = f_1(1),\ g(n+1) = \max\,[g(n), f_r(p)] + 1 \text{ for } r \leqslant n,\ p \leqslant n;$$

note that $g(n) > n$, and, for $p > n$, $g(p) > f_n(p)$. Thus the property $V(n)$ is decidable, where $V(n)$ holds if, and only if, n is a *value* of g, i.e. for some m, $g(m) = n$.

Consider the (decidable) ordering $a \prec b$ where $a \prec b$ if, and only if, (i) $\rightarrow V(a)\ \&\ \rightarrow V(b)\ \&\ a < b$, or (ii) $\rightarrow V(a)\ \&\ V(b)$, or (iii) $V(a)\ \&\ V(b)\ \&\ b < a$.

Every decreasing sequence $f_n(1), f_n(2), \ldots$ is finite, but $g(1), g(2), \ldots$ is an infinite decreasing sequence.

If the sequence of functions $f_n(m)$ is not recursively enumerable, the definition of the ordering still goes through, but the relation $a < b$ need not be systematically decidable.

Thus, *given an enumerable sequence of functions we have constructed an ordering which is a well-ordering for sequences defined by these functions, but not for arbitrary sequences.*

CYBERNETICS AND MENTAL FUNCTIONING *

R. Thomson and W. Sluckin

It is now approximately a decade since Ashby in this country and then Wiener and others in America began to write on the subject since christened ' cybernetics ' ; and the interest in this field and its ramifications does not appear to be abating. J. O. Wisdom has pointed out that at the foundations of this whole field of study is a single and essentially testable hypothesis, namely, that negative feed-back mechanisms underlie the working of the central nervous system.[1] Much of what is written on the subject consists either in bringing forward evidence for the hypothesis or discussing corollaries that can be drawn from it. But discussion ranges also into psychiatry, various fields of psychology and even into sociology; it embraces the subject of ' men and machines ', and bears upon the traditional questions of philosophy of mind and metaphysics.

What is the point of the non-hypothetical discussions of cybernetics ? What is the contribution of cybernetic thought to those social sciences upon which it impinges ? How does it obtrude in philosophical discussion ? Our purpose is to classify and interpret those discussions which seem to go beyond the cybernetic hypothesis and its supporting evidence. The analysis will lead us to the not unfamiliar problem of the role of analogies, models and ' metaphysical climate ' in constructive, scientific thinking.

1 *Problem-solving*

In mechanical systems embodying negative feed-back, the action of the system is reduced when its result (the machine output) approaches a certain level, which is an equilibrium level for the system. Should an extraneous cause upset this state of balance, the activity will increase or decrease as required to re-establish the equilibrium. This can occur because information regarding the deviation from equilibrium, that is the amount of error, is fed back to that part of the system which controls the level of activity, regulating the latter automatically.

* Received 21. x. 52

[1] J. O. Wisdom, ' The hypothesis of cybernetics ', this *Journal*, 1951, 2, 1-24

In an organism, the concept of homeostasis refers to the equilibrium level of activity. If a homeostatic condition is upset, activity increases or decreases as necessary until equilibrium is restored. This happens because ' information ' regarding the amount of unbalance is fed back by the autonomic nervous system to the mechanisms which regulate bodily activity associated with the particular bodily condition. The regulation is automatic in that the amount of unbalance determines the level of activity directed towards the restoration of balance.

Homeostatic condition is a goal for the activity of an organism. But, in fact, any goal of animal or human activity constitutes an equilibrium level analogous to the equilibrium of a mechanical system. If a person's goal at a particular moment is, say, to stand up, then standing up represents from the standpoint of the central nervous system a state of equilibrium, even if so only for a fleeting moment. In general, in goal-seeking behaviour, there is negative feed-back of the results of activity controlling subsequent level of activity.

The attempt to reach a goal may be regarded as representing a problem to the organism. When the goal is actually reached, the problem has been solved. Often the solution of a problem is attained after a series of attempts. If the problem is a simple one, there is no evidence of any discontinuity of activity such as is present when a series of discrete attempts is made to reach a goal. Nevertheless, we suppose that there is always some kind of trial and error, feed-back of error, further trial and so on, though often all this fuses into continuous activity.

It may be said that in each case of error, information regarding the success or lack of success of the trial is ' fed back ' to influence further activity. Lack of success leads to further attempts which may land the animal or person near the goal. Complete success at attaining the goal, i.e. solving the problem, stops further modifying activity ; for the equilibrium has been reached.

So we may talk, if we wish, about problem-solving behaviour in terms of feed-back. It may not be particularly helpful to describe trial-and-error behaviour as being based upon negative feed-back. At least, it is interesting to note that the simple concept of negative feed-back is applicable not only to mechanical and electrical automatic regulation devices, to the autonomic and the central nervous systems, but also to problem-solving behaviour. When we apply the concept to behaviour, we are using it, perhaps, not in its strict

technical sense ; we are employing a picture or a mechanical model. We are not really making a hypothesis about the nature of problem-solving behaviour. Rather, we are using a new model for describing what had previously been more or less adequately described without it.

The possibility of employing this sort of model for thinking about thought and behaviour is given fuller attention later.

2 *Learning*

By learning we mean acquiring skills, habits, and ideas. Learning goes on as we solve problems or even as we fail to solve them. Clearly, much learning takes place by trial and error. Solving a problem makes it more likely that, in future, success in a similar situation will be achieved more easily and speedily.

In learning a manual skill we try persistently until all the incorrect moves have been eliminated. In acquiring a habit less and less effort is wasted on eliminating incipient wrong moves until the habit becomes ingrained. Obtaining knowledge, too, consists essentially in developing correct reactions to mental clues. It may be said that negative feed-back ensures that unless success is complete we go on trying.

At the turn of the century E. L. Thorndike formulated the Law of Effect. Briefly this states that successful features of behaviour are stamped in, and the unsuccessful ones eliminated. In other words, each success modifies subsequent activity of the learner in the direction of increasing the probability of selecting such a step again, or each failure leads to a lesser probability of subsequently attempting a similar step. In the simplest case, a failure will make it certain that the wrong move will not be repeated, while a success will make it certain that the learner will make the correct move again on a future occasion.

It is theoretically not difficult to design a machine that will learn in this sense of the word. Mr I. P. Howard of the Department of Psychology, University of Durham, has actually constructed a 'mechanical rat' which will run *any* maze (provided the width of the lanes is within certain limits). It will, of course, make mistakes during the first trial, though it will always successfully complete the run. On subsequent attempts it will make no mistakes. It will have 'learned' not to do so by 'experience'.

To make mechanical learning more like trial-and-error learning of living creatures, a mechanism could be made to be capable of initially responding to a situation in a number of ways; and which response it will make can be made to be a matter of chance. Once, however, the response labelled correct has been made, the chances of it being made again on a subsequent occasion will be a little greater than before, and so on, until after some time the probability of a correct response will become virtually a certainty. Such a mechanism will 'learn by experience' just as an organism learns by experience.

Overt trial-and-error is, of course, not the only kind of learning. We have not so far mentioned conditioning, or learning by association, or learning by insight. In recent years it has become increasingly clear to psychologists that the differences between the various kinds of learning are not of a fundamental nature. 'The "kind" of learning which the experimenter finds depends on the nature of the problem which he sets, on what he is looking for, on what aspects of the subject's activity he chooses to observe.'[1]

It may be plausibly maintained that in every kind of learning incorrect responses, whether overt or incipient, are eliminated, and correct ones are stamped in. And this makes learning a process which, basically, can be imitated by a machine.

The most primitive view of learning is associated with what Popper calls 'the bucket theory of mind'—a primitive mechanical model. More advanced views of learning are associated with the theory that knowledge manifests itself as modifications of reaction to external environment. This involves a more elaborate mechanical (or electronic) model.

Learning is associated with intelligence. This has been described by reference to behaviour only, as 'the property of reacting on a basis of probability as determined by the individual organism's incomplete sensory samples of the environment'.[2] It is interesting that it has been possible to design an electrical circuit, or to put it more picturesquely, a 'hypothetical nervous system' in conformity with data from behaviour experiments, which will exhibit intelligent behaviour in this sense.

[1] W. N. Kellogg in *Methods of Psychology*, ed. T. G. Andrews, New York, 1948

[2] H. E. Coburn, 'The Brain Analogy', *Psychol. Rev.*, 1951, 58, 155

3 *Thinking*

Psychologists tend to regard thinking as essentially 'what occurs in experience when an organism, human or animal, meets, recognises, and solves a problem'.[1] Problem-solving behaviour may be overt and observable, but it may also be covert when it occurs in thought. In other words, problem-solving is implicit in thinking.

Before the advent of cybernetics, Craik had argued that 'thought models, or parallels, reality'. Its essential feature is symbolism; 'this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation'.[2]

The succession of our thoughts, clearly, does not consist of complete solutions of problems. Many thought processes are interrupted as new questions turn up, and the threads are picked up again later. We may regard a thought process as complete when a problem has been solved. In this sense some thought processes go on for years, while others are begun and ended within one fleeting moment.

Just as in overt problem-solving, so in thinking there is a continuous feed-back of information regarding the failure of successive trial answers or guesses. This feed-back keeps the thought process going, regulating mental activity and maintaining tension. Mental equilibrium, as we might call it, is attained only when the task of answering the initial question has been successfully completed.

The analysis of the methods of the empirical sciences known as the hypothetico-deductive system suggests that progress in science is like the progress of a thought process. A question or problem demanding explanation is posed and a first hypothetical answer or explanation is put forward. This is a 'trial' answer; its success is measured by the amount of agreement between it and deductions from it, and empirical tests. Feed-back of information regarding the results of the tests of hypotheses keeps the process going. New hypotheses are made when old ones have been falsified. The working of a negative feed-back mechanism models such an account of the methods of the natural and social sciences.

Considerations of the nature of thought, coupled with rapid progress in the construction of digital computers and analogue

[1] G. Humphrey, *Thinking*, London, 1951

[2] K. J. W. Craik, *The Nature of Explanation*, Cambridge, 1943

machines, have led to discussions in which it is claimed by some and denied by others that robots may be described as capable of thinking. It is also both claimed and denied that machines exhibit purposeful behaviour.

Before examining these disputes we must determine in what senses the concepts of thought and behaviour are applied to human beings. Once we say what we mean when we talk about human thinking and human behaviour we can decide whether or not robots can be described as thinking or behaving, in precisely the same sense in which humans are so described. If these concepts are being applied to robots in a new sense, what are the differences between descriptions of human mentality and descriptions of robot mentality? What criteria do we use in deciding whether to apply or withhold these descriptions; what sort of distinctions are we making when we use these concepts in ordinary situations?

Contemporary philosophy presents interpretations of 'thinking' which contribute to this discussion. Ryle,[1] for example, distinguishes the following uses:

(i) 'think' is often used as a synonym for 'believe', 'suppose', 'doubt' (I don't think), 'understand', 'imagine', etc.; in these uses it is possible to think all sorts of nonsense; such beliefs are frequently induced or suggested by such irrational means as propaganda and advertising;

(ii) 'think' often means 'be in a particular frame of mind'; footballers, rock-climbers and wrestlers may all think what they are doing, in this usage, while actively engaged on their respective pursuits;

(iii) 'think' may mean work out or puzzle out a problem (cogitate);

(iv) 'thinking' sometimes means stating a conclusion or using a conclusion intelligently (physicians applying their theories to a case).

Ryle makes a further distinction between 'thought' in the sense of 'work' (X is thinking it out) and in the sense of 'results' (what X thought out; the conclusions reached).

It is claimed that robots think in senses (iii) and (iv). They have 'thoughts' both in the sense of 'working out problems' and in the sense of giving 'results' for publication and criticism. They perform mathematical operations and they solve chess problems. They reach conclusions and may be able to apply conclusions intelligently to new

[1] G. Ryle, *The Concept of Mind*, London, 1949

situations (Ashby's homeostat).[1] What is entailed in our ordinary discourse when we talk about working out problems and when we talk about using conclusions ?

Consider the working out of problems : puzzling, pondering (compare (iii) above). What we mean by these terms cannot be summarised in a neat formula or definition. Such terms are highly metaphorical (' weighing ', ' turning over in one's mind ', ' racking ', ' unravelling '). In this process there are soliloquies, calculations and miscalculations, asking and answering specific questions, debating, cross-examining oneself, etc. What form these operations take cannot be formularised. Three people given the same problem will set to work differently in the ' reflecting ', ' pondering ' or ' thinking it out ' : the using of paper and pen, writing and re-writing notes, and perhaps using slide-rules, log tables, instruments, reading and copying parts of books, articles—all go to make up this process.

In theory the worker could utter thoughts aloud instead of saying them to himself, and thus his moves and counter-moves while working towards the completion of the task could be described by a B.B.C. commentator. Ryle insists that ' thinking ' is not some peculiar shadow-process which goes on behind the overt performances of the thinker and which is only describable in special terms denoting ' mental acts '. ' Thinking ' is a performance involving the exercise of capacities to do certain operations in a certain manner. What kind of a performance, involving what particular capacities and executed in what characteristic manner depends upon the problem which the person attempts to solve. The terms ' thinking ', ' thought ', ' reflection ', etc., are general concepts denoting the more concrete descriptive concepts which indicate specific performances : these latter words refer to concrete cases where X asks himself the question ' p ' and responds by making suitable replies to himself, works out problems with the aid of mathematical tables, and does other similar intellectual jobs.

A person described as ' thinking ' may be exhibiting any one of a wide range of behaviour-patterns. To analyse the concept of ' thinking ' in this sense would be to write a treatise investigating the logical and epistemological peculiarities of every branch of human enquiry, as well as an infinite number of commonsense situations.

[1] W. R. Ashby, ' Design for a Brain ', *Electronic Engineering*, 1948, **20**, 379–383. See also W. R. Ashby in *Perspectives in Neuropsychiatry*, ed. D. Richter, London, 1950 ; and *Design for a Brain*, London, 1952

Consider now the products of thinking (compare (iv) above). We exhibit the products of thought when we have reached conclusions and are able to apply them in appropriate situations. ' What X thinks ' can be analysed in the form : if situations of kind S arise, X reacts by stating certain conclusions, by presenting certain arguments, by applying certain canons of criticism in the moves he makes, etc. He exhibits certain dispositions to use conclusions in particular ways.

This characterising of the use of ' thinking ' might be said to classify ' thinking ' as a disposition, namely, the capacity to do certain things, in a certain frame of mind, in response to an appropriate stimulus or problem. Part of this ' doing of certain things ' is symbol-using (words, diagrams, images ' in the mind's eye ', asking and answering questions in soliloquy) ; and part is executing actions (scribbling, referring to books, tables, slide-rule operating. Such dispositional behaviour tends towards a specific kind of result, and might be described as ' goal-directed '.

The result or ' goal-state ' is either the statement of conclusions, or the disposition to apply conclusions in appropriate situations. Thought in any sense is a variety of dispositional behaviour.

Since thinking is analysed as a variety of dispositional behaviour it is not difficult to find analogies between what a human being does who is described as thinking and what a robot does when it works out problems and produces conclusions. The behavioural criteria, whereby we decide whether or not to describe a man as ' thinking ', tempt us to allow robots to pass the test for ' thinking '. Against this, objections can be raised which attempt to show that cognitive concepts have implications which refer to the wider context within which human thought as distinct from robot ' thought ' occurs. Descriptions apply to human beings in this wider context which do not fit robots, descriptions which characterise the inner life of the human being or which refer to complex social relations.

At the same time it can be shown, perhaps, that cognitive concepts *can* be interpreted, in certain uses, in purely behavioural terms ; in these significant cases it is difficult to withhold the descriptions from machines, especially when these have been constructed for the purpose of doing jobs which were formerly only possible if done by human intelligence. Here there are arguments for and against the extension of the use of a family of terms (which were formerly applied exclusively to human agents) to electronic artefacts. When it comes to

describing any phenomena in general terms we find that the language at our disposal is ambiguous.

Is the dispute over the question 'Do robots think or do they not?' merely an indirect discussion on the advisability of extending the meaning of a well-established but ambiguous set of terms? Apart from formulating the hypothesis which claims to be able to explain how it is that human beings have certain capacities, is the discourse of cybernetics a means of exhibiting verbal confusions arising out of the systematic ambiguity of the concepts of intelligence? Before answering this question we must face the related problem 'Do robots exhibit purposeful behaviour?'

4 *Purposeful Behaviour*

When we describe a human being as behaving purposefully we sometimes imply that his action is a suitable subject for ethical appraisal. It does not appear that the concept of purpose can be applied to robots in this sense. We cannot blame a robot for the results of its operations or praise it for meritorious action, except in metaphor. Robots do not form their purposes as humans do since their purposes are decided for them by their inventors or operators; nor do robots have ethical sentiments or attitudes; they do not function, as human beings do, in a context of social relationships which evoke moral sentiments and the language appropriate to the expression of such sentiments. At the present stage of machine development there does not seem to be any robot situation analogous to human situations which give rise to the evocation of moral judgments. It is not profitable to discuss the question whether or not robots exhibit purpose in the sense in which human purposes are discussed in morals: at present they cannot even be described in terms which ascribe morality to their behaviour.

When, however, we describe human beings as behaving purposefully we do not always imply that their purposes are liable to moral appraisal. We may imply merely that the actions classified as purposeful are the deliberate exercise of a capacity, ability or skill to do something or other. Such dispositional actions can be the subject of moral appraisal, but there are instances where this is not the case. The dispute over the problem 'Do machines exhibit purpose?' may be interpreted in the form of the question 'Do robots exhibit dispositional behaviour?'—for example, do they

acquire skills and make intelligent adaptations to changing situations in their surroundings? What, then, do we mean when we ascribe dispositional behaviour to a human being?

When a person is described as having the capacity to do something we mean that he can exercise his skill if he applies himself to the task; he knows how to do it and it requires only the appropriate effort for the result to be achieved. When the agent does actually apply himself and succeeds, we regard him as having behaved purposefully. To apply the concept of purpose to human activity is to assert that a performance produces a specific result, that the purposeful activity is the effect of learning, and that the actions promoting the result are carried out in a characteristic manner or—to use Ryle's phrase—'a characteristic frame of mind'. To describe a frame of mind is simply to describe the more minute characteristics of the performance; when we describe a deliberate act we say what the result was and what were the most obvious steps leading to the result: in addition to these we might mention the deliberate frame of mind in which the marksman aimed (his fine adjustment of eye, sights, target, his holding his breath when pressing the trigger, etc.).

To use the word 'purpose' is either to mark out the action as intelligent or dispositional or else to characterise the frame of mind in which the action was performed. In each case we are saying something about describable features of overt behaviour and the conditions under which such behaviour takes place.

It may be noted that 'purpose' is an ambiguous concept.

(i) It serves to classify human behaviour as deliberate.

(ii) There is another sense in which any process which tends to produce a particular effect is 'purposeful'. Machines exhibit purpose in this sense of the word, for example, 'the purpose of this device in the slot-machine is the detection and rejection of imperfect coins'. Many non-human activities—the behaviour of insects or electric circuits, in fact any dynamic process which tends to repeat a particular result, exhibit 'purpose' in this sense.

Involuntary actions are contrasted with purposeful actions. We have no control over these responses, they are actions we cannot help doing. It is not reasonable to expect the agent to exercise any check or control over reflexes, instincts, impulses, ingrained habits, accidents. Although some of these actions are 'purposeful' in sense (ii) above, we mark them off from all higher order purposes (sense (i) above). We *might* also exclude from the class of purposeful actions

internal compulsions (the effects of nervous disorder, hypnosis, drugs, intoxication) and external compulsions (actions which we say we were 'forced' to do under threat of violence or because of compelling circumstances). In the latter categories there arise difficult borderline cases which puzzle judges and jurymen, but the general distinction is firmly established in ordinary usage.

In addition to these uses in which it is applied to humans, the verb 'behave' is extended to apply to any activity exhibited by a living thing and to the operations of machines and, by analogy, to almost any dynamic system.

This elementary analysis of the general distinctions implied by our use of the notion of 'behaviour' in relation to human beings enables us to make some observations on the dispute arising from the question 'do robots exhibit purposeful behaviour?' In one sense of the phrase 'purposeful behaviour', it is absurd to dispute the assertion 'robots display purposeful behaviour'. 'Behave' applies to the working of machines as well as to the reactions of human beings, and 'purpose' merely means 'brings about a specific result as the end-state of a series of movements'.

More than this, however, is claimed on behalf of the robots. They are regarded as exhibiting dispositional behaviour in the same sense in which we speak of human intelligent action. When faced with this claim it is easy to fall into a state of conflict, to argue for and against the claim. The reason for this situation is revealed when we examine the criteria governing our use of the concept of 'purpose' as it applies to dispositions.

When we attempt to say what we mean by 'behave purposefully' in the sense 'exercise capacity or skill', we find ourselves talking about overt performances and the conditions which promote them; the analysis is given in behavioural terms; the criteria for 'purposeful action' are behavioural. This is not surprising since these criteria are readily detected and described. When we examine the sort of criteria used in distinguishing what a human being does, when described as exhibiting purposeful behaviour of this kind, we notice that there is a close analogy between these features of human action and what a robot does when it performs a similar task.

However rigorously we refine our characterisation of the criteria which we employ to define this type of human action, we find it difficult to make clear-cut differences between our characterisations of human dispositions and our characterisations of robot simulations of

human dispositions: the same sort of descriptions seem to apply to each. It is tempting to suggest that the robot qualifies for the application of concepts of intelligence, and in so doing qualifies as 'exhibiting dispositional purposeful behaviour'.

Discussions which attempt to show that robots behave purposefully take the following form: (*a*) a description of what robots do and how they do it; (*b*) some sort of analysis or interpretation of what we mean when we apply concepts of intelligence to human beings; (*c*) illustrations which suggest that the criteria for applying concepts of intelligence to human beings allow us to extend our use of these words to robot performances. Those who object to the conclusion that robots behave purposefully try to find overtones of meaning in our descriptions of human dispositions, which do not fit robot performances (for example, attempt to show that to describe a person as behaving purposefully in the sense under discussion implies that the person is activated by certain motives or emotions or interests; that reference to personality and temperament are involved, etc.).

A thorough investigation of ordinary usage of dispositional words would reveal only that when we use such terms we do not make precise distinctions or imply anything in the nature of an exhaustive account of the actions denoted. We say something significant about dispositional behaviour when we use this family of concepts carefully, but not very much is said. If it can be shown that there is little difficulty in extending the use of such words to describe robots as well as human beings, this merely shows that the concepts in question are vague as regards scope and have a wider range of possible uses than had been recognised.

Whether or not one ought to extend the use of these words to cover robots as well as human beings is a matter of epistemological nicety.

Are these disputes over 'thought' and 'purpose' in machines concerned chiefly with the working out of the proper uses to which certain expressions can be put? Recent arguments between Turing and Mays,[1] and between Spilsbury and MacKay [2] show that the disputants are aware of this aspect of the discussion. These discussions nevertheless exhibit misleading trends. The problem is stated as if it were a question of fact whether or not machines think; a question

[1] A. M. Turing, 'Computing Machinery and Intelligence', *Mind*, 1950, **59**, 433; W. Mays, 'Can Machines Think?', *Philosophy*, 1952, **27**, 148–162.

[2] R. T. Spilsbury, D. M. MacKay, 'Mentality in Machines', *Proc. Arist. Soc.*, 1952, Suppl. vol. 26

demanding a 'Yes' or a 'No' answer. Facts are relevant, but it is known what they are (what robots can or even might do and how they do it). What we are puzzled about is whether or not robots may be sensibly described in certain ways, given the facts and given the conventions of our language. The most unsatisfactory feature of discussions to date has been the tendency of disputants to argue as if there were only one use of a key concept which stands as *the* proper or essential use (for example, MacKay's definition of 'behaviour', p. 62, and 'purpose', p. 63).

There has been no recognition of the philosophical doctrine that to analyse the meaning of a concept requires the exhibition of the entire range of its different, although logically related, uses; and the demonstration that any one of these distinct uses can only have a 'proper' application on a specific occasion of use, the specific occasion determining the precise inflexion of meaning which the concept has in that particular context.

In some contexts concepts can be said to have strictly definable meanings. In others they cannot be so readily limited in possible meaning; they exhibit greater flexibility of use. Hence there are some arguments about how words ought to be extended or limited in meaning which resemble some kinds of argument over borderline cases in morals and aesthetics; after rational and dispassionate dispute there remains no agreement; the use of argument only succeeds in making precise the fine distinctions which divide the disputants. The worry over the application of concepts of mentality to robots would seem to be an example of this latter kind of 'endless' dispute. Once the facts have been fully examined, and the logical analysis of the concepts has been executed, we are left with the possibility of deciding by the toss of a coin. Which way we decide for extension or for constriction of meaning—does not matter, so long as we know what we are implying, and avoid paradox and confusion.

To carry out the discussion in terms of only one significant use (or range of uses), however general in application this use may be, is to invite dispute over assertions which express only half-truths (namely to imply that robots think in every possible use of the term 'think', or on the contrary that they can never be described as 'thinking' in any possible use of the term). Once a reasonably adequate analysis has been accomplished we see that there is nothing to dispute; or rather that the question 'Do robots think and display

purposeful behaviour' is a misleading question, encouraging us to take sides when our task is only properly done when we demonstrate that this is a case where there are two sides.

This is not to say that the discussions referred to above do not succeed in bringing out the use of concepts of mentality: they do; but they do not succeed in bringing out clearly subtle inflexions of meaning and the criteria governing these. Although cybernetics gives rise to such exercises in analytical philosophy there is no reason to conclude that the provocation supplied by the puzzle 'can machines think?' is necessary for the attainment of such philosophical insight; or that anything has emerged from these discussions which constitutes an advance in this kind of inquiry.

5 *Analogies, Models and Metaphysics*

In addition to exercises in philosophical analysis, cybernetics provokes a different kind of philosophical thinking. In cybernetic and related writings there is an attempt to interpret commonsense descriptions of human mentality in a way which makes such descriptions apply equally well to men or machines. In effect, these interpretations raise the ghosts of old metaphysical disputes. In what respects are the writings of cyberneticians metaphysical? When it is maintained that robots think, those who make this claim make also, by implication, another claim. They are able to offer a new interpretation of what it is that a man does when he is described as 'thinking'. Thus it might be said that when a man thinks he is displaying goal-directed activity which is controlled by negative feed-back mechanisms throughout his nervous system.

This is something more than a statement of the cybernetic hypothesis. The analogy of a robot solving problems, which human beings solve when they think, is used to provide a model for what a man is doing who is described as 'thinking'. In statements of this kind our ordinary discourse for talking about human mentality is 'corrected' or 'improved' by relating what we mean when we describe X as 'thinking' to the hypothesis (which purports to show how X has this capacity to think). The 'co-ordinating analogy' expressed in such statements relates a commonsense description to the hypothesis which explains whatever the description describes.

This analogy is further extended to serve a different purpose from that of expressing a co-ordination of science with commonsense.

The analogy between human beings and robots is made the basis for a model which has wider implications. This further analogy emphasises that, whatever differences there may be between human beings and robots, they resemble each other in one respect ; each may be regarded as a self-regulating system, controlled by intricate mechanisms which are 'built-in' to the system. Every action and reaction of the system can be explained wholly by reference to the operations of these mechanisms. This is the essential analogy which is suggested by the cybernetic hypothesis. This analogy is made to work as a model in terms of which all forms of human mentality can be interpreted. The purpose of the model is to stand as a symbol for a set of presuppositions. What these underlying assumptions are, is made clear by F. C. S. Northrop [1] in an article in which he discusses the views of Rosenblueth, Wiener and Bigelow, together with those of W. S. McCulloch and W. Pitts. These assumptions, which the model of man as a robot controlled by negative feedback mechanisms symbolises, are simply a reformulation of the traditional metaphysical doctrine known as Scientific Materialism.

Northrop quotes from cybernetic writings, showing how these discussions offer 'solutions' to traditional metaphysical problems. It is claimed that the 'Mind-Body' problem discussed by metaphysicians since Descartes is solved by the use of cybernetic models. Objections against earlier attempts of Materialists to explain consciousness, thinking, memory, recognition of form, etc., in terms of the mechanics of the nervous system are met by using the analogy. The older model of machines exhibiting what is now called the 'open-cycle' type of mechanism is replaced by the model of the 'closed-cycle' type. The latter arrangement can be either mechanical or electronic. The electronic model provides a greater complexity of response and adaptation. The analogy implies the belief that the human nervous system will be explicable in terms of the principles governing 'closed-cycle' devices. Electronic engineering supplies a new model for thinking about psychological and biological processes.

The traditional problem of how purely mechanical systems can exhibit purposeful activity without the intervention of some non-mechanical (teleological) system of causes is also solved in favour of scientific materialism by the application of electronic-mechanical models. Only random activity is nonpurposeful. Purposeful

[1] F. C. S. Northrop, 'The neurological and behaviouristic psychological basis of the ordering of society by means of ideas', *Science*, 1948, **107**, 411-416

activities can be shown to be classifiable into two kinds: (*a*) teleological purposes are controlled by negative feedback mechanisms: signals from the goal alter the activity after it has been initiated, so that the active system achieves its goal by mechanical reactions controlled from the goal-object or goal-state; (*b*) non-teleological purposes are those which proceed towards a goal-state but which have no negative feedback controls. Purposes are not excluded from mechanical systems as such but only from random-behaviour mechanisms. Teleological systems can be wholly mechanical systems. The objections of Vitalists in biological metaphysics and Idealists in psychological metaphysics to the doctrines of Scientific Materialism are therefore undermined.

By implication, although not explicitly, the answer is supplied to the whole system of disputes over the nature of causation, the relation of mind to body, the nature of consciousness, the nature of substance, the nature of voluntary action, etc., which metaphysicians debated in the disputes between Vitalist and Mechanist, Idealist and Materialist. This is, for example, the case when an eminent neurologist suggests that neuropsychoses are due to structural damage in the brain, that functional psychoses are due to improper voltages at nerve cells (which function like battery-operated magnetic relays), that neuroses occur when some of the negative feed-back loops become positive or regenerative, while the so-called psychodynamic theories are nonsense.[1] This is again the case when cyberneticians make statements with a view to discrediting isomorphism.[2] Isomorphism, a species of psychophysical parallelism, is a metaphysical adjunct of Gestalt psychology, and cannot be falsified or confirmed by empirical evidence of any kind. Examples of metaphysical preoccupations in the writings of Wiener and his associates could easily be multiplied.

It would be unfair to dismiss the metaphysical discussions in cybernetics as unimportant. The analogy between the functioning of living creatures and the simulation of it by electronic artefacts has led to the formulation of a scientific hypothesis. Evidence is accumulating in favour of this hypothesis suggesting that human thought and purpose are guided and controlled by the mechanical adjustments of the nervous system to signals coming from the

[1] W. S. McCulloch, 'The Brain as a Computing Machine', *Electrical Engineering* 1949, **68**, 492–497

[2] W. Pitts and W. S. McCulloch, 'How do we know universals?', *Bull. Math. Biophys.*, 1947, **9**, 127

environment, much as robots are guided and controlled in their simulations of human behaviour. Metaphysical statements in cybernetics serve two distinct functions: they express, through the device of the co-ordinating analogy, the most general implications of the cybernetic hypothesis; they also express the analogy of 'Man as something resembling a robot' in order to proclaim confidence in the cybernetic hypothesis as the most rational and comprehensive starting point for attacking the problem of explaining human thought and purpose.

Nevertheless, much of the excited tone in cybernetic writings appears when purely metaphysical problems are resurrected. There is no harm done by the reformulation of the case for Scientific Materialism. It is doubtful if much of positive value is achieved by reconstructing this or any other metaphysical system. Apart from giving biologists and psychologists the emotional satisfaction of a *credo*, it is unlikely that the discussion of familiar traditional metaphysical problems, in a refurbished form, will have the stimulating effect which a new metaphysics can have: suggesting new hypotheses for testing, promoting improved methods of philosophical analysis and leading to new forms of ordinary speech for talking about the world in a fresh way. The principal worth of cybernetics is to be found in those writings which confine themselves to what is at the root of the discussion: the negative feed-back hypothesis of neurophysiology.

Departments of Psychology and of Philosophy
University of Durham

SEMANTIC INFORMATION *

YEHOSHUA BAR-HILLEL and RUDOLF CARNAP

I

THE Mathematical Theory of Communication, often referred to also as Theory (of Transmission) of Information, as practised nowadays, is not interested in the content of the symbols whose information it measures. The measures, as defined, for instance, by Shannon,[1] have nothing to do with what these symbols symbolise, but only with the frequency of their occurrence.[2] The probabilities which occur in the definientia of the definitions of the various concepts in Communication Theory are just these frequencies, absolute or relative, sometimes perhaps estimates of these frequencies.

This deliberate restriction of the scope of Statistical Communication Theory was of great heuristic value and enabled this theory to reach important results in a short time. Unfortunately, however, it often turned out that impatient scientists in various fields applied the terminology and the theorems of Communication Theory to fields in which the term 'information' was used, presystematically, in a semantic sense, that is, one involving contents or designata of symbols,

* Received 23. x. 52. This is a slightly revised version of a paper read before the Symposium on Applications of Communication Theory, London, 26th September 1952. A more detailed and systematic treatment of the same topic appears simultaneously in *Technical Report* No. 247 of the Research Laboratory of Electronics, Massachusetts Institute of Technology, 1953.

[1] See, e.g., C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949. E. Colin Cherry's 'A History of the Theory of Information', *Proceedings of the Institution of Electrical Engineers*, 1951, **98**, part iii, pp. 383-393, gives an excellent account of the development of this theory and contains also an extensive bibliography up to 1950.

[2] A notable exception to this general trend of communication engineers is D. M. MacKay, who recognised as early as 1948 that the concept of information treated in Communication Theory, which he proposes now to call 'selective information' should be supplemented by a concept of 'scientific information'. It seems, however, that this concept does not coincide with what we call here 'semantic information'. A clarification of the exact relationship will be undertaken elsewhere.

The clearest presentation of MacKay's ideas may be found in his contribution to the *Transactions of the Eighth Conference on Cybernetics*, 'In Search of Basic Symbols', New York, 1952, pp. 181-235, where an earlier paper of his, 'The Nomenclature of Information Theory', is also reprinted.

or even in a pragmatic sense, that is, one involving the users of these symbols. There can be no doubt that the clarification of these concepts of information is a very important task. However, the definitions of information and amount of information given in present Communication Theory do not constitute a solution of this task. To transfer these definitions to the fields in which those semantic or pragmatic concepts are used, may at best have some heuristic stimulating value but at worst be absolutely misleading.

In the following, the outlines of a Theory of Semantic Information will be presented. The contents of the symbols will be decisively involved in the definition of the basic concepts of this theory and an application of these concepts and of the theorems concerning them to fields involving semantics thereby warranted. But precaution will still have to be taken not to apply prematurely these concepts and theorems to fields like psychology and other social sciences, in which users of symbols play an essential role. It is expected, however, that the semantic concept of information will serve as a better approximation for some future explanation of a psychological concept of information than the concept dealt with in Communication Theory.

2

The fundamental concepts of the theory of semantic information can be defined in a straightforward way on the basis of the theory of inductive probability that has been recently developed by one of us.[1] Unfortunately, the space at our disposal does not permit us to develop the full terminological background on which our presentation is based. We have to refer once and for all to the extensive presentation of this background given in [Prob.] or to the more concise one given in [Cont.]. (In the Appendix, an even more concise summary is offered for the convenience of the reader.) Let us state only that what follows refer to a fixed language system L_n^π, by which we mean, approximately, an applied first-order functional semantical system with n individuals, say a_1, a_2 . . . , a_n, and π primitive properties, say P_1, P_2, . . . , P_π. A disjunction which, for each of the πn atomic statements, contains either this statement or its negation (but not both) as a component, will be called a *content-element*. The content-elements are the weakest factual statements of L_n^π inasmuch as the

[1] R. Carnap, *Logical Foundations of Probability*, University of Chicago Press, 1950, cited hereafter as [Prob.] and *The Continuum of Inductive Methods*, University of Chicago Press, 1952, cited hereafter as [Cont.].

only factual statement L-implied by a content-element is this content-element itself. One of the 64 content-elements in L_3^2, for instance, is

$$P_1a_1 v \sim P_2a_1 v \sim P_1a_2 v P_2a_2 v P_1a_3 v P_2a_3.$$

The class of all content-elements L-implied by any statement i (in L_n^π) is called the *content* of this statement and denoted by 'Cont (i)'. It can be easily verified that the content of any atomic statement contains exactly half of all content-elements, that of an L-true statement none, and that of an L-false statement all of them. The last property may look slightly artificial but is no more so than the use of, say, the null-set in set-theory.

We offer Cont (i) as an explicatum for the ordinary concept 'the information conveyed by the statement i', taken in its semantic sense. We have no time to show at length that Cont (i) is an adequate explicatum. But it can be immediately verified that it fulfils at least the condition that Cont (i) includes Cont (j) if i L-implies j. This condition should be regarded as a necessary, though certainly not sufficient, condition of adequacy of any proposed explanation of the mentioned concept.

Since Cont (i) is equal to the class of the negations of the state-descriptions contained in the range of $\sim$ i, the properties of Cont (i) can be easily derived from the properties of the concept 'range of i' which has been treated at length in [Prob.]. We shall say nothing more here.

It is often important not only to know what is the information conveyed by some statement but also to attach a measure to this information. We need not start afresh looking for appropriate measure-functions ranging over contents since measure-functions over ranges have been extensively discussed in [Prob.]. For each of the latter m-functions, as they are called in [Prob.], a corresponding content-measure-function is defined simply by

$$\text{cont (i)} = \text{m} (\sim \text{i}).$$

cont (i) (read: the content-measure of i) is offered as one (not *the*) explicatum of the ordinary concept 'amount of information conveyed by i' in its semantic sense. Among the most important properties of cont (i), immediately derivable from the corresponding properties of m(i) treated in [Prob.] we have

$$0 \leqq \text{cont (i)} \leqq 1,$$

where the extremes are reserved for L-true and L-false statements, respectively, and

$$\text{cont}\ (i\,.\,j) = \text{cont}\ (i) + \text{cont}\ (j) - \text{cont}\ (i\vee j).$$

From the last theorem follow immediately :

$$\text{cont}\ (i\,.\,j) \leqq \text{cont}\ (i) + \text{cont}\ (j),$$

and the interesting additivity theorem :

$$\text{cont}\ (i\,.\,j) = \text{cont}\ (i) + \text{cont}\ (j) \text{ if and only if i and j are L-disjunct.}$$

Here, however, an inconsistency in the intuitions of many of us becomes apparent. Though it is indeed, after some reflection, quite plausible that the content of a conjunction should be equal to the sum of the contents of its components if and only if these components are L-disjunct or content-exclusive, in other words, if they have no factual consequences in common, it is also plausible, without much reflection, that the content of the conjunction of two basic statements, say 'P_1a_1' and '$\sim P_2a_3$' should be equal to the sum of the contents of these statements since they are independent, and this not only in the weak deductive sense of this term, but even in the much stronger sense of initial irrelevance. But no two basic statements with different predicates are L-disjunct, since they have their disjunction, which is a factual statement, as a common consequence. Our intuitions here, as in so many other cases, are in conflict and it is best to solve this conflict by assuming that there is not *one* explicandum 'amount of information' but at least two, for one of which cont is a suitable explicatum, whereas the explicatum for the other has still to be found.

So far we have dealt with the information conveyed by some statement separately. At times, however, we are as much, or even more, interested in the information conveyed by a statement j in excess to that conveyed by some other statement i or a class of statements. We therefore define the concepts 'content of j relative to i' and 'content-measure of j relative to i' by

$$\text{Cont}\ (j/i) = \text{Cont}\ (i\,.\,j) - \text{Cont}\ (i)$$

and

$$\text{cont}\ (j/i) = \text{cont}\ (i\,.\,j) - \text{cont}\ (i)$$

respectively. (Notice that the '—' in the first of these definitions is the symbol of class-difference, in the second that of ordinary numerical difference.) The maximum value of cont (j/i) is obviously cont (j) and this value is obtained if and only if i and j are L-disjunct. The

minimum value of cont (j/i) is 0 and this value is obtained if, and only if, i L-implies j. Of special interest is that

$$\text{cont}\ (j/t) = \text{cont}\ (j),$$

where t stands for any L-true statement (any ' tautology '), since this allows us to define cont (j) in terms of cont (j/i), thereby reversing the definition procedure followed by us before. Even more interesting is that

$$\text{cont}\ (j/i) = \text{cont}\ (i \supset j),$$

from which it follows that, given i, j conveys no more additional information than i ⊃ j, by itself a much weaker statement.

If we stipulate now, for the second explicatum of ' amounts of information ', that all basic statements shall convey the same amount of information, and this independently of whether these statements appear alone or as components in some non-contradictory conjunction, and if we stipulate, in addition, for the purpose of normalisation, that the amount of information conveyed by a basic statement shall be 1, it can easily be seen, along well-known lines of computation, that these stipulations are fulfilled if we define this second function, to be called ' measure of information ' and denoted by ' inf ', as

$$\text{inf}(i) = \text{Log}\ \frac{1}{1\text{-cont}\ (i)}$$

(where ' Log ' stands for $\log_2$), from which we obtain, by simple substitution,

$$\text{inf}(i) = \text{Log}\ \frac{1}{m\ (i)} = -\ \text{Log}\ m\ (i).$$

The last equation is analogous to a definition of amount of information in Communication Theory but with inductive probability instead of statistical probability.

Among the various theorems regarding inf let us mention

$$0 \leqq \text{inf}(i) \leqq$$

and the theorem of additivity, which, however, involves now a quite different condition.

$\text{inf}(i\,.\,j) = \text{inf}(i) + \text{inf}(j)$ if, and only if, i is initially irrelevant to j (with respect to that m-function on which inf is based).

Sometimes $\text{inf}(i\,.\,j)$ is greater than $\text{inf}(i) + \text{inf}(j)$. If anyone should find this strange that might be due to the fact that he has subconsciously switched to some other explicatum, such as cont, for which this can indeed not happen.

Another theorem of great importance deals with inf(j/i). It states that

$$\inf(j/i) = \text{Log}\,\frac{1}{c(j, i)} = -\,\text{Log}\,c(j, i)$$

where c(j , i) is the degree of confirmation of (the hypothesis) j on (the evidence) i, defined in [Prob.] as $\frac{m(i\,.\,j)}{m(i)}$.

The statistical correlate of inf has found a large field of application in communication engineering. Neither cont nor its statistical correlate have found any useful application so far. It is, however, to be expected that the facet of the amount of information which is measured by cont, should find its fields of application too, especially so since cont is a mathematically simpler function of m than inf.[1]

3

Among the various m-functions on which cont and inf may be based, there are two groups of special importance. The first group consists of just one member, to be designated here by 'm_D'—its symbol in [Prob.] is '$m^\dagger$' — : the second group has infinitely many members denoted collectively by 'm_I'. m_D assigns to each content-element the same value. This makes the computations with this function especially easy, in general, and the preference given to it understandable. It suffers, however, from the great disadvantage that it does not allow us, roughly speaking, to learn from experience. 'P_1a_4', for instance, will have a c_D-value of $\frac{1}{2}$, on no evidence at all or, in other words, on the tautological evidence, and the same c_D-value on the evidence '$P_1a_1 . P_1a_2 . P_1a_3$'. In spite of this defect, there are situations in which m_D , c_D, and the information-functions based upon them may be of importance. Situations in which we intend to use only deductive reasoning are of this type, hence the subscript 'D' for 'deductive'.

In those situations in which inductive reasoning is to be applied, only such m-functions may be used which allow us to learn from experience, in other words, which fulfil the Requirement of Instantial

[1] Indeed, in a paper by John G. Kemeny and Paul Oppenheim, 'Degree of Factual Support', *Philosophy of Science*, 1952, **19**, 288-306 (published after the present paper was read in London), a concept of Strength of a statement was used whose definition corresponds closely to that of our cont_D.

Relevance.[1] These are the functions m_I; the ' I ' stands for ' inductive '.

All the theorems that hold for cont and inf in general hold, of course, also for $cont_D$ and inf_D and all the $cont_I$ and inf_I functions. For these more specific functions, however, additional theorems can be proven. For lack of space, this will not be done here. Let us only remark that certain inconsistencies in our intuitive requirements with regard to information-functions may be due to a subconscious switching from D-type functions to I-type functions and vice versa.

4

Situations often arise in which we do not know whether a certain event has occurred or will occur, but only that exactly one event out of a class of mutually exclusive events has occurred or will occur. The statements describing these events convey each a certain amount of information on the available evidence. It makes therefore good sense to ask for some average of the amount of information conveyed by these statements. If these statements refer to future events, one talks about the amount of information that may be expected to be conveyed, on the average. In [Prob.] it is shown that in many similar situations the c-mean estimate of the function in question will be a satisfactory measure for this expected value. Confining ourselves here, for the sake of easy comparability with prevailing Communication Theory, to inf and using ' exhaustive system ' to denote a class of statements of the above-mentioned character, we define *the (c-mean) estimate of the measure of information conveyed by (the members of the exhaustive system) H on (the evidence) e*, in symbols : est (inf, H, e), as follows :

$$\text{est}(\text{inf}, H, e) = \sum_{p=1}^{n} c(h_p, e) \times \text{inf}(h_p/e).$$

From this definition and from a prior theorem on $\text{inf}(h_p/e)$ the theorem

$$\text{est}(\text{inf}, H, e) = -\sum_{p} c(h_p, e) \times \text{Log}\, c(h_p, e)$$

immediately follows. The communicational correlate of this theorem is well known. We see no reason, so far, to attach any special significance to the formal similarity of its right side to certain entropy-type expressions in statistical thermodynamics.

[1] See R. Carnap, ' On the Comparative Concept of Confirmation ', this *Journal*, 1953, **3**, p. 314

To give a simple illustration: If on the basis of available evidence, mainly prior observations, the c-value of the hypothesis, h_1, 'There will be warm weather in London on the 23rd of September 1953',[1] is $\frac{1}{2}$, the c-value of h_2, 'There will be temperate weather . . .', is $\frac{1}{4}$, and the c-value of h_3, 'There will be cold weather . . .', is $\frac{1}{4}$, then

$$\text{est}(\text{inf}, H, e) = -\sum_p c(h_p, e) \times \text{Log}\, c(h_p, e)$$
$$= \tfrac{1}{2} \times 1 + \tfrac{1}{4} \times 2 + \tfrac{1}{4} \times 2 = 1{\cdot}5$$

(where $H = \{h_1, h_2, h_3\}$).

If $H = \{h_1 \ldots h_n\}$ and $K = \{k_1 \ldots k_m\}$ are (deductively) independent exhaustive systems (on e) (i.e. no $h_p \,.\, k_q$ is L-false (on e)), then $H \,.\, K$, defined as $\{h_1 \,.\, k_1, h_1 \,.\, k_2 \ldots h_1 \,.\, k_m, h_2 \,.\, k_1 \ldots h_n \,.\, k_m\}$ is exhaustive too (on e), hence

$$\text{est}(\text{inf}, H \,.\, K, e) = \sum_{p=1}^{n} \sum_{q=1}^{m} c(h_p \,.\, k_q, e) \times \text{inf}(h_p \,.\, k_q/e).$$

We have

$$\text{est}(\text{inf}, H \,.\, K, e) \leqq \text{est}(\text{inf}, H, e) + \text{est}(\text{inf}, K, e),$$

with equality holding if, and only if, the h's and the k's are mutually irrelevant.

If the statement k is added to our evidence, *the posterior estimate of the measure of information conveyed by H on e and k* will, in general, be different from *the prior estimate of the measure of information conveyed by H on e alone.* This difference is often of great importance and will therefore receive a special name, *amount of specification of H through k on e* and be denoted by 'sp (inf, H, k, e)'. The formal definition is

$$\text{sp}(\text{inf}, H, k, e) = \text{est}(\text{inf}, H, e) - \text{est}(\text{inf}, H, e \,.\, k).$$

It is easy to see that $\text{sp}(\text{inf}, H, k, e) = 0$ if (but not only if) k is irrelevant to the h's on e. sp may have positive and negative values with its maximum obviously equal to the prior estimate itself. This value will be obtained when $e \,.\, k$ L-implies some h_p. In this case, H is completely specified through k (on e).

Let, to continue our previous illustration, k_1 be a certain report of weather-instrument-readings. Let $c(k_1, e \,.\, h_1) = \frac{1}{4}$, $c(k_1, e \,.\, h_2) = c(k_1, e \,.\, h_3) = \frac{1}{2}$. The following diagram will help to visualise the situation:

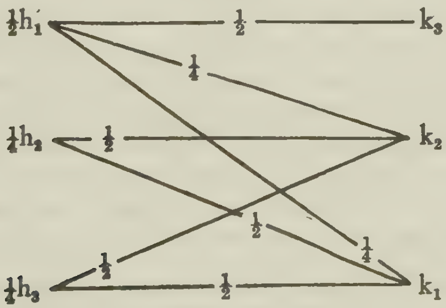
[1] The formulation of this statement exceeds already the potentialities of the language-systems envisaged here. However, this is of no importance in this connection.

It is easy to compute, or to read from the diagram, that $c(h_1, e \,.\, k_1) = c(h_2, e \,.\, k_1) = c(h_3, e \,.\, k_1) = \frac{1}{3}$. Hence $\text{est}(\text{inf}, H, e \,.\, k_1) = \text{Log } 3 = 1{\cdot}585$ and $\text{sp}(\text{inf}, H, k_1, e) = -0{\cdot}085$.

Situations often arise in which the event stated in k has not yet occurred, or, at least, in which it is not known whether it has occurred but in which it is known that either it or some other event belonging to a certain class of events will occur or has occurred. In such circumstances, it makes sense to ask for some average of the posterior estimate of the measure of information conveyed by H on e and (some member of) K (the exhaustive system of the k's). We are led to *the (c-mean) estimate of this posterior estimate* denoted by ' est(inf, H/K, e) ' and defined as

$$\text{est}(\text{inf}, H/K, e) = \sum_{q=1}^{m} c(k_q, e) \times \text{est}(\text{inf}, H, e \,.\, k_q).$$

Let us complete our illustration in the following diagram :



Then $\text{est}(\text{inf}, H/K, e) = \sum_{q} c(k_q\, , e) \times \text{est}(\text{inf}, H, e \,.\, k_q)$

$$= \tfrac{3}{8} \times \text{Log } 3 + \tfrac{3}{8} \times \text{Log } 3 = 1{\cdot}189.$$

The estimate of the amount of specification of H through K on e is, of course, equal to the difference between the prior estimate of the

measure of information conveyed by H on e and the estimate of the posterior estimate of the measure of information conveyed by H on e and K, in symbols :

$$sp(\text{inf}, H, K, e) = est(\text{inf}, H, e) - est(\text{inf}, H/K, e).$$

In our example, sp(inf, H, K, e) = 1·5 — 1·189 = 0·311.

Let us mention only three theorems in this connection, the communicational correlates of which are well-known :

$$est(\text{inf}, H/K, e) = est(\text{inf}, H . K, e) - est(\text{inf}, K, e),$$
$$sp(\text{inf}, H, K, e) = sp(\text{inf}, K, H, e),$$
$$sp(\text{inf}, H, K, e) \geqq 0.$$

5

Lack of space prevents us from going any deeper into the significance of the concepts and theorems indicated in the last section. It seems that the theory of semantic information might be fruitfully applied in various fields, for instance in the Theory of Design of Experiments [1] and in Test Theory.[2]

In view of the many misunderstandings and misapplications, in which Communication Theory has been involved, it would be desirable to undertake a clarification of its foundations. So would be a comparison between the theory outlined here and Communication Theory. These two tasks will be undertaken elsewhere by one of the authors (B.-H.).

Appendix

The language systems dealt with in this paper contain a finite number of *individual constants* which stand for *individuals* (things, events, or positions) and a finite number of *primitive one-place predicates* which designate primitive properties of the individuals. In an *atomic statement* e.g., 'P_1a_1' ('the individual a_1 has the property P_1'), a primitive property is asserted to hold for an individual. Atomic statements and statements formed out of one or more of them with the help of the customary connectives of negation,

[1] Indeed, R. A. Fisher defined in *The Design of Experiments*, Edinburgh and London, 1935 (in a less developed form already in papers dating back to 1922), a concept of Amount of Information which is, however, only distantly related to that developed here. Fisher's concept is certainly not a communicational one, but it is, like the communicational one, of a statistical, and not of a semantical, nature.

[2] Recent work done by Lee J. Cronbach at the University of Illinois seems to point in this direction. See, e.g., his preliminary report 'A Generalised Psychometric Theory Based on Information Measure', mimeographed, March 1952.

'$\sim$' ('not'), of disjunction, 'v' ('or'), of conjunction, '.' ('and'), of (material) implication, '$\supset$' ('if . . . then'), and of (material) equivalence, '$\equiv$' (if and only if), are *molecular statements*. All atomic statements and their negations are *basic statements*. It is known that with the help of these tools numerical statements can be formed. Hence absolute frequencies (cardinal numbers of classes or properties) and relative frequencies can be expressed in them (but not measurable quantities like length and mass).

Any sentence is either *L-true* (logically true, analytic, e.g. '$P_1a_1 \vee \sim P_1a_1$') or *L-false* (logically false, self-contradictory, e.g. '$P_1a_1 . \sim P_1a_1$') or *factual* (logically indeterminate, synthetic, e.g. '$P_1a_1 \vee \sim P_2P_3$'). Logical relations can be defined, e.g. 'The statement i *L-implies* the statement j' for 'i $\supset$ j is L-true', 'i is *L-equivalent* to j' for 'i $\equiv$ j is L-true' 'i is *L-disjunct* with j' for 'ivj is L-true'.

A *state-description* is a conjunction containing as components for every atomic statement either this statement or its negation, but not both, and no other statements. Thus a state-description completely describes a possible state of the universe in question. For any statement j of the system the class of those state-descriptions in which j holds, i.e. each of which L-implies j, is called the *range* of j. The range of j is null if, and only if, j is L-false ; in any other case j is L-equivalent to the disjunction of the state-descriptions in its range.

Research Laboratory of Electronics,
Massachusetts Institute of Technology,
Cambridge, Mass., U.S.A.

Yehoshua Bar-Hillel

Department of Philosophy,
University of Chicago,
Chicago, Ill., U.S.A.

Rudolf Carnap

NOTES AND COMMENTS

Order, Organisation and Entropy

THE application of the Second Law of Thermodynamics to biological systems is a subject of current interest to both physicists and biologists.[1] One important point which is involved in the biological applications of the Second Law is the relationship between the concept of organisation, as used by the biologist, and the concepts of order and entropy, as used by the physicist. It is generally assumed that an increase in the biological organisation of a system, such as occurs during the embryological stages of an organism, is equivalent to, or at least closely connected with, a decrease in the entropy of the system, entropy being the physicist's measure of disorder. It is the purpose of this note to discuss the relationships between organisation and entropy and to point out that the usual assumption, as stated above, need not be valid in all cases. (Since one cannot compute entropies for biological systems, the examples given will be chosen from simpler chemical systems, but the principles involved are not thereby affected.)

Our first point is that thermodynamic states of a system which vary widely in the degree of their organisation may nevertheless have the same entropy. This assertion is readily demonstrated by an examination of the entropy at the absolute zero of temperature. According to the Nernst theorem, or Third Law of Thermodynamics, there can be no entropy difference between different thermodynamic states of a system at absolute zero.[2] If we take as our system suitable numbers of atoms of carbon, hydrogen, oxygen, nitrogen and sulphur, then we can have at absolute zero a number of thermodynamic states of this system which differ radically in the degree of their organisation. Thus, the system can consist of separate crystals of the pure elements or of separate crystals of such simple molecules as H_2O, CO_2, SO_2, etc., or of amino acid crystals, or, finally, it can consist of a single protein crystal. Since all of these modes of organisation of the atoms correspond to thermodynamic states of the same system, it follows.

[1] Erwin Schrödinger, *What is Life*, Cambridge, 1946, Ch. 6, 7
Harold F. Blum, *Time's Arrow and Evolution*, Princeton, New Jersey, 1951

[2] R. H. Fowler and E. A. Guggenheim, *Statistical Thermodynamics*, Cambridge, 1939, Ch. 5
Erwin Schrödinger, *Statistical Thermodynamics*, Cambridge, 1948, Ch. 3

from the Third Law that they must all have the same entropy at absolute zero.

Our second point is that it is possible to have two thermodynamic states of a system such that the one which is the more highly organised has, at the same time, the larger entropy, i.e. the greater disorder, in the physicist's sense of the word disorder. This assertion may be demonstrated by consideration of the following example. Let us compare the entropy, S_1, of two moles of hydrogen chloride gas in a container of volume V at room temperature with the entropy, S_2, of a mixture of one mole of hydrogen gas and one mole of chlorine gas in a container of the same volume V, also at room temperature. The standard methods of statistical thermodynamics[1] give for the difference of entropies, $S_1 - S_2$, the following result

$$S_1 - S_2 = R\left\{\ln \frac{\theta_{H_2}\theta_{Cl_2}}{(\theta_{HCl})^2} + \frac{3}{2}\ln \frac{(m_{HCl})^2}{m_{H_2}m_{Cl_2}}\right\}$$

where R is the gas constant, θ_{H_2}, θ_{Cl_2} and θ_{HCl} are the characteristic rotational temperatures of the three molecular species and m_{H_2}, m_{Cl_2} and m_{HCl} are the molecular weights of the three species. Using the numerical values, $\theta_{H_2} = 85{\cdot}0\ °K$, $\theta_{Cl_2} = 0{\cdot}35\ °K$, $\theta_{HCl} = 14{\cdot}9\ °K$ and $m_{H_2} = 2$, $m_{Cl_2} = 71$, $m_{HCl} = 36{\cdot}5$ we obtain

$$S_1 - S_2 = 1{\cdot}35\ R$$

It is seen that $S_1 - S_2$ is certainly positive in this example, i.e. the more highly organised hydrogen chloride gas has an entropy greater than that of the mixture of hydrogen and chlorine gases. This example does not, of course, imply that the more highly organised system always has higher entropy, but only shows that there are cases in which the converse proposition is false.

If one analyses the meaning given to the entropy concept in statistical mechanics, it is not surprising that entropy and organisation are as unrelated as the above examples demonstrate. One can say that the entropy is low when the number of energy states, in which there is a non-negligible probability of finding the system, is small. The entropy does not depend explicitly upon the nature of the wave functions for the system which are associated with these energy states. It is in the wave function, however, that the structure of the system is reflected, and it is this structure which is associated with the concept of organisation. Thus, in the first example discussed above, the

[1] Fowler and Guggenheim, *op. cit.* Ch. 3

equality of the entropies at absolute zero for the different state of organisation meant that in each case the system was sure to be found in its lowest energy state. The wave functions associated with these states would be very different depending upon whether the system consisted, for example, of elementary crystals or of a single protein crystal.

We conclude that the degree of order in a system which is measured by the entropy (low entropy corresponding to high order) is not the same thing as the degree of organisation of the system in the sense used by the biologist.

The work discussed above was suggested by a paper of Needham's.[1] Our conclusions, though arrived at by different argument, are, we believe, in general agreement with Needham's.

The author would like to thank Professor Erwin Schrödinger for helpful criticism.

MARTIN J. KLEIN

School of Theoretical Physics
Institute for Advanced Studies
64-65 Merrion Square, Dublin

(National Research Fellow in Physics, on leave of absence from Case Institute of Technology, Cleveland, Ohio, U.S.A.)

Light Signal Kinematics

THE following episode is instructive on the relation of formal deductive systems to physical theories and the ease with which fundamental mistakes or misunderstandings can arise.

By 1935 four workers[2] had developed partly-formalised deductive theories of relativistic (Minkowski) kinematics, seeking to use only light signals and the relation of temporal succession. *Each of these four*[3] appeared in certain passages to be maintaining a *physical*

[1] Joseph Needham, *Science and Society*, New York, 1942, vol. 6, pp. 352-375

[2] A. A. Robb, *A Theory of Time and Space*, Cambridge, 1914, and later books; C. Carathéodory, *Sitzb. Preuss. Akad. Berlin*, 1924, p. 12; H. Reichenbach, *Axiomatik der Relativistischer Raum-Zeit Lehre*, Braunschweig, 1924; *Zeit. f. Phys.*, 1925, **34**, 32; E. A. Milne, *Relativity, Gravitation, and World Structure*, Oxford, 1935; *Kinematic Relativity*, Oxford, 1948, and elsewhere

[3] Carathéodory retains the surplus transformations, which he regards as 'trivial'. Reichenbach published a correction (*loc. cit.*, 1925). In 1929 I called Robb's attention to this in relation to his own work. In 1935 I lent Milne the *Axiomatik*, mentioning the mistake and correction. He replied (25th March 1935) that he had not yet

proposition which is incorrect,[1] namely that using only light signals and temporal succession (without either a clock-process or a rod acceptable to the physicist) it is possible to construct ordinary measures of length and time. A physicist using only light signals cannot discriminate inertial systems from these subjected to arbitrary 4-D similarity transformations. The system of 'resting' mass-points which can so be identified may be arbitrarily expanding and/or contracting relatively to a rod, and these superfluous transformations can only be eliminated by using a rod or a clock. In a formal system restrictive conditions at infinity can serve instead, a fact of no operational significance.

L. L. WHYTE

The Time of Psychology and of Physics

THE issue of this *Journal* for May 1952, containing Dr Scott Blair's interesting comments upon my paper [2] has only just come into my hands. Despite the lapse of time I should like to reply briefly to his points.

First, Dr Scott Blair criticises my suggestion of time-uncertainties of the order of the duration of a specious present.[3] I agree that on any definition of time-uncertainty which makes it a function of the smallest space-interval which can be sharply defined time-uncertainties of the order contemplated by me cannot be expected. But this conception of time-uncertainty is only a special application of Heisenberg's general uncertainty principle to the particular case of *elementary* particles such as electrons, protons, and mesons. It is appropriate to a theory of elementary indeterminancy such as Furth's,[4] but has no

reached a firm opinion on the point. In some passages Milne uses atomic clocks ; in others he seems to suggest that he can obtain the Lorentz transformations without clocks. Cf. A. G. Walker's fully formalised system, *Proc. Roy. Soc. Edin.* 1946, **62,** 319.

[1] H. Weyl, *Deutsche Literaturz.*, 1924 (col. 2122 ; brilliantly lucid review of *Axiomatik*) ; A. S. Eddington, *Nature*, 1935, **135,** 635 ; J. L. Synge, *Nature*, 1936, **138,** 28 ; G. C. McVittie, *Proc. Roy Soc. Edin.*, 1942, **61,** 210. In a broader context, cf. G. J. Whitrow, *Phil. Mag.*, 1946, **37,** 469.

[2] This *Journal*, 1951, **2,** 122, 177

[3] which has been given as half a second or so ; vide E. B. Hebb, *The Organization of Behaviour*, New York, 1949, p. 74

[4] *Nature*, 1950, **166,** 30

application to macroscopic structures such as the human nervous system.

It is clear that the general uncertainty relation $\Delta E \,.\, \Delta t \sim h$ permits of time-uncertainties of any order as we fix the energy level more and more sharply. Moreover the converse of this effect can actually be observed in certain phenomena : e.g. in molecular predissociation, where restricting the time-interval allowed for transitions specially sharply results in broadening of the corresponding spectral lines.

In terms of the relation $\Delta E \,.\, \Delta t \sim h$ there *can be no upper limit* to the size of the time-uncertainty as we approach completely sharp definition of energy level. In a completely stationary state with a theoretically perfectly determinate energy there is infinite time-uncertainty. In fact the time is so indefinite that, as Schrödinger says, ' where anything happens we are not facing pure energy states '.[1]

So much for the possibility of uncertainties in ' time intervals of the order of a few seconds ', which possibility is a clear consequence of quantum mechanics. As to the ' multiplicative effect ' of resonance on the quantum of least uncertainty, the matter is more difficult and no acceptable theory exists at present. The effect has, however, been noted in qualitative discussions ; [2] and as I hope to indicate in a separate publication the basis for a quantitative theory I will not discuss this aspect further here.

Secondly, I agree with Dr Scott Blair as to the importance of the work of Rhine and others such as Soal, which suggests a spread of compresence extending beyond the normal limits and even into the ' apparent ' future. I say ' apparent ' because in the most recent development [3] of my two-fold time theory I put forward the conception of time as a type of complex variable in which what I have called phase-time is the ' pure imaginary ' component. In phase-time there are relations of precedence which make it proper to say of two event-phases (*a*) and (*b*) that : if (*a*) precedes (*b*), and (*a*) is now ' present ' (*b*) is ' quasi-future ' ; even though *both* are comprésent together in a world-wide moment of becoming which is itself a four-dimensional cross-section of the five-fold world. I believe apparent precogitation is to be explained in terms of an abnormal, or statistically infrequent, type of orientation of the experient with respect to space-(phase) time : so that he apprehends contents of the total

[1] This *Journal*, 1952, **3**, 122

[2] e.g. Rice and Teller, *Structure of Matter*, New York, 1949, p. 274

[3] To be published shortly

momentary world-spread of compresence which are quasi-future w.r.t. the specious presents of most 'normal' human observers.

Thirdly, I am not sure that I agree with Dr Scott Blair's account of Dunne's Infinite Regress. As Broad has pointed out McTaggert anticipated Dunne's argument for a Regress, but McTaggert drew the exactly opposite conclusion: namely, that it proved the unreality of Time. If Broad's analysis is correct both men made the same *basic* error of confusing the extensional with the transitory aspect of time.

Fourthly, I entirely agree with Dr Scott Blair that multiplying a real time by an algebraic square root of minus one does not abolish the difference between time and space in the sense that Minkowski seems to have had in mind. This was a fallacy of which Eddington also was guilty when he talked of an imaginary time-co-ordinate as equivalent to a real space-like co-ordinate.[1]

On my view, however, there is a *logical* difference between time and space which is more fundamental than the ontological differences which Dr Scott Blair mentions.

For the ordinal generating relation for a linear *time* point series is the dyadic one of 'precedence'; whereas that for a linear *spatial* point series is the triadic one of 'betweenness'.

Finally, apropos of the theory of time-scales, I would merely say that the reference to Milne in my paper was obscure through being over-brief. What I had in mind was that although Milne himself apparently regarded his time-scales as applying within the same dimension a careful analysis of his theory suggests that this is logically impossible. For if we accept relativity then we must accept the principle that an arbitrary but consistent change of scale in the co-ordinates used to describe an objective phenomenon cannot alter the 'objective nature' of that phenomenon since this phrase simply refers to invariance under such 'gauge transformations'. Now on Milne's account there are instances where a regraduation of clocks results in a different *objective phenomenon.* The most notorious case being that of the famous 'red-shift'[2] in the light from distant nebulae. For a change of time-scale here from 't' time to 'τ' time converts a Doppler effect into a degradation of frequency in a photon with age.

[1] *Fundamental Theory*, Cambridge, 1946, p. 125

[2] See his *Modern Cosmology and the Christian Idea of God*, Oxford, 1952, where a modified account of the red-shift is given.

Thus in this case a regraduation changes a reversible into an irreversible phenomenon. The only way out is to regard the two time-scales as different dimensional co-ordinates : then it will make sense to say that the red-shift is a Doppler effect taken with respect to *t*-time (the time dimension in regard to which the universe is expanding), but is a frequency degradation with respect to the other time.

H. A. C. Dobbs

REVIEWS

Metallurgy in Antiquity, R. J. Forbes (E. J. Brill, Leiden, 1950. Pp. 489. 19 guilders.)

JUST as Greek mathematics has a permanent value, so many of the techniques of metal working developed in antiquity encompass an understanding or intuition of the properties of ores and metals that was to advance no further until the later Middle Ages or the Renaissance. The Greek mathematicians, it has been said, are rather like ' fellows of another college ' and the smiths of antiquity have much in common with the metallurgical craftsmen known to Biringuccio and Agricola. This perpetuation of crude but effective processes for hundreds of years and the diversity of the subject matter which ranges from early texts to archaeological finds, emphasises the need for a critical appreciation of historical method in any account of metallurgy in antiquity.

Professor Forbes' book is admirable in this respect. The first four chapters cover a synopsis of early metallurgy and mining techniques, the development of the craft of the smith, and his social and sacred status. Then follow six chapters on the various metals in turn : gold ; silver and lead; tin; antimony and arsenic; zinc and brass; copper; and finally iron. In these, the discussion ranges from the location and type of ore worked in antiquity, the smelting and refining methods used, the finishing techniques adopted such as hammering, heat-quenching, and casting methods, to the distribution of the metals among the various cultures and the trade routes which were followed. Very wide use has been made here of the minutiae of archaeological, linguistic and technological scholarship. Extensive bibliographies are appended to each chapter, containing in the case of those dealing with the individual metals an average of 140 references. There are 98 illustrations including many maps of mining localities and mineral deposits.

One of the first historical problems is the choice of terms. ' Bronze Age ' and ' Iron Age ' have obvious limitations, for in the Middle East they overlap almost inextricably, and in Africa the use of iron preceded that of bronze. A division based on the smith's growing knowledge of the metals, independent of their kind, is far more fundamental. First native metals were used as if they were stones, then their unique properties were appreciated over a period which saw the widespread use of copper, gold, silver and meteoric iron. This was followed by a period in which ores were smelted ; the composition of a resulting metal alloy was the most

significant factor, as in the case of bronze and brass. Then finally came the smelting of iron ores when the processing of the metal was the most important factor, different treatments producing soft iron, wrought iron, cast iron or steel. It is against such a background that the development of mining methods, tools and metal technology can best be appreciated and related to the growing knowledge of still more mineral species. On this basis, after a palaeolithic and neolithic age up to 3500 B.C., Forbes distinguishes a predynastic age 3500–3000 B.C., two metal ages 3000–2200 B.C. and 2200–1200 B.C., an early and late iron age 1200–500 B.C. and 500–50 B.C., followed by the period of the Roman Empire 50 B.C.–A.D. 300.

The most difficult historical problem is the evaluation and correlation of the various sources of knowledge. Even in the later period when accounts like those of Pliny and Strabo are available, the writing may be elegant but the metallurgy faulty or incomplete. For the earlier there is an obvious paucity of data ; and so the occasional Egyptian or Sumerian inscription dealing with metallurgical topics has great interest, significance and charm. Mines in Sinai were worked by the Egyptians and stelae set up in the region tell the details of mining expeditions. But all such written sources need great care in the interpretation of the words : often copper or iron are used when the real meaning is bronze or steel.

There are three important aspects of the archaeological evidence : first, how the style and form of the object is related to the properties of the chief metals in use at the time ; secondly, whether chemical or metallurgical analysis enables the source and perhaps the origin of the metal or ore to be decided ; and thirdly, whether the extraction or smelting procedures and finishing processes can be inferred. The use of iron in the Near East illustrates this first aspect. It appears first in jewelry, amulets and statues, then as an ornament for bronze implements, followed by its use for the working part with bronze as the ornamentation, and finally it is used for the entire object which now has its form dictated by the particular properties of iron. Chemical analysis gives unequivocal evidence on the source of the metal in several instances. Native gold is invariably an alloy, average values being Californian—88·4 per cent., Australian—95 per cent. and Japanese—60–92 per cent. It is therefore certain that the gold used in the ancient Near East was mainly the native alloy and that refining was not practised until the later phase of the period. Silver obtained from the native gold-silver alloy electrum retains much gold in contrast to that obtained from lead ores which only has about 0·3 per cent. Native copper is extremely pure, 99·9 per cent., whereas that from ores would only reach about 98 per cent. with different metallic impurities. Meteoric iron, which was the kind first used, contains on the average 7·5 per cent. of nickel and some cobalt, unlike smelted iron. The nature of the extraction, smelting and finishing techniques used in antiquity has been the subject of much

speculation. Some of the problems have been solved by the researches of Coghlan, who showed that ordinary fires without an air blast could only bring about the reduction of copper ores and that higher temperatures, such as those obtained in pottery furnaces, were required for melting the metal preparatory to casting. An interesting corollary to this is that whereas iron ores can be reduced to the metal at lower temperatures than copper ores, the latter yield a lump of metal, but the former a spongy 'bloom' containing much slag which has to be repeatedly heated and hammered to obtain a mass of metal. The non-metallic appearance of this 'bloom' and its recalcitrance may well be the reason for the use of copper, and hence bronze, before iron.

Other surprising conclusions follow from metallurgical and mineralogical insight as a few examples will show. Ignorance of the metallic nature of one component of an alloy did not prevent its use; for instance, the presence of tin in bronze when it was first used, was unsuspected, likewise the presence of zinc in brass. Brass is a very striking instance, for metallic zinc was unknown since it sublimes below the reduction temperature of its ore and so can only be prepared in a distillation apparatus, with effectively a reducing atmosphere, which was developed much later. For this reason the history of brass follows the geology of zinc, because brass was made by the treatment of metallic copper with zinc ores. The superiority of the iron ores of Noricum depended on manganese impurities yielding a 'natural steel' independent of any knowledge of the carbonising-quenching process for toughening wrought iron. Such an accidental occurrence contributed much to the development of the Hallstatt civilisation. Lastly, there is good evidence that silver production on a significant scale was a by-product in the manufacture of lead, and that the early smelting of iron ore, as distinct from the use of meteoric iron, arose out of gold extraction, because the dense and rich ore, magnetite, is often associated with gold in the river gravels.

It is natural to ask the question, what form science took in this preclassical period, to what extent it had emerged beyond a mass and medley of phantasy and uncorrelated observations. In its fashion it was a model of precision and practical classification, a careful description and nomenclature was sufficient, and the place of everything in the cosmos and the role it was to play was clearly defined. The period of explanation and the understanding of structure and mechanism was yet to come. The lists of minerals and drugs, which the Sumerians compiled, demonstrate this beautifully. The minerals were arranged according to outward appearance and properties such as hardness and colour, and according to the metal each was thought to contain. A cumulative nomenclature was developed based on a group name and a determinative, closely related in principle to that now used in organic chemistry. The first stage of a modern science,

classification, had in fact been achieved. However, even in the Greek classical period, when logical patterns were being sought, there is an absence of one particular correlation that might well have been expected. In spite of the interest in mathematics, particularly geometry, no attempt was made to classify minerals according to their crystalline form, a procedure which was to be so fruitful in modern times.

The discussion of these and many other themes, with a wealth of factual detail make this book invaluable to archaeologists and historians of science and technology.

PHILIP GEORGE

A History of Science, Technology, and Philosophy in the sixteenth and seventeenth centuries, A. Wolf. New edition prepared by Douglas McKie. (George Allen & Unwin Ltd., London, 1950. Pp. xxviii + 692. 42s.)

A History of Science, Technology, and Philosophy in the eighteenth Century. Second edition revised by D. McKie (George Allen & Unwin Ltd., London, 1952. Pp. 814. 60s.)

THESE two volumes were first published in 1935 and 1938, respectively, and are now republished with corrections and additions to the bibliographies by Dr McKie. Together they comprise an invaluable source of information on most aspects of the history of science in the period of the 'Scientific Revolution'. The topics covered range over the internal history of the various natural sciences: mathematics, mechanics, astronomy, light, sound, electricity and magnetism, meteorology, chemistry, geology, botany, zoology, anatomy, medicine; there are chapters on the scientific academies, scientific instruments, exploration and cartography, technological problems of agriculture, manufacturing, building, mining, engineering, power-machinery, and transport; and at the end of each volume come accounts of contemporary contributions to psychology and the social sciences, and brief summaries of the views of contemporary philosophers, taking account of their relation to natural science. As this recitation will suggest, these volumes are a work rather of compilation than of synthesis: they do not give so good a picture of the position of science in the civilisation of the time as Preserved Smith's *History of Modern Culture*, which covers the same period; and they entirely miss the opportunity to show how science produced a systematic revolution in thought, as seen especially in the development of ideas on method. But nowhere else will so wide a range of accurate, detailed information about sixteenth-, seventeenth- and eighteenth-century science

be found in so convenient a form. Both volumes are very well and fully illustrated.

A major criticism of the presentation given here of the ' scientific revolution ' is that it repeats the old-fashioned view that no contribution was made to scientific thought in the Middle Ages ; the attempt, in the first chapter of the first volume, to characterise the ' medieval mentality ' is in fact altogether naïve and ignorant. The writings of Duhem are nowhere mentioned ; nor are the more recent publications of Gilson, C. B. Boyer, Koyré and others who have continued Duhem's investigation of the relationship between medieval and seventeenth-century thought ; Dugas' admirable *Histoire de la Mécanique* appeared too late to be included. Readers interested in the general problem of opinion about the ' scientific revolution ' should consult Wallace K. Ferguson's *Renaissance in Historical Thought* (1948).

Prescinding from this criticism this new edition of Professor Wolf's volumes may be thoroughly recommended.

A. C. Crombie

Design for a Brain. W. Ross Ashby (Chapman & Hall Ltd., London 1952. Pp. ix + 259. 36s.)

A fair and just assessment of the value of this book requires its discussion on two distinct planes, those respectively of physiology and philosophy. The author propounds an interesting theory about the way in which one aspect of the central nervous system in the higher vertebrae works, claiming it to be comprehensive enough to explain, in the words of the dust cover ' the origin of the nervous system's power of being self-organising and adaptive ' and to answer the very big philosophical question ' how can the brain, as a mechanism, also be intelligent ? '

Dr Ross Ashby's method differs from the one most commonly employed in physiological research. When a physiologist wants to investigate the working of the nervous system he normally examines samples of this system, dissects them, traces out all their ramifications, ascertains what physical and chemical changes they undergo while they are working, conducts sundry experiments on them, and thus acquires bit by bit a knowledge of the function of each component part and the way in which it works. Dr. Ross Ashby's approach is more theoretical. He has constructed a machine, which he calls a homeostat, the mechanism of which makes it behave in one respect somewhat like an animal or a human being adapting itself to its environment.

The theory takes as its starting point the well-known fact that in living tissues, as in many engineering devices, such as servo mechanisms and other so-called closed loop systems, automatic control is effected with the help of a principle called feed-back. Devices using this principle have means by which some quantity such as temperature, acidity, velocity, alignment, is kept constant under changing conditions. A deviation from a specified value of such a quantity sets mechanisms in motion that reduce the deviation.

In the homeostat the specified quantity is the position of a pointer, and the machine is so designed that its performance in maintaining the pointer within certain limits can be upset when other parts of the machine are disturbed. When this happens the pointer does not return but continues to deviate and the machine is said to have become unstable. In an ordinary servo mechanism one could restore stability by making suitable adjustments to some of the circuit constants. In the homeostat these adjustments are made automatically. While deviating beyond certain limits the pointer keeps scanning a number of successive circuit constants arranged at random. When connection is made to constants that restore stability, the pointer returns to its specified position and no more change to the constants can occur.

The homeostat may thus be regarded as having become proof against the particular disturbance to which it was subjected. Its performance, to this extent, is analogous to that of an animal that has adapted its pattern of behaviour to a changed environment. But only to this extent ; all that the homeostat does is to maintain a pointer within specified limits and such a limited performance can hardly be called truly comparable to a pattern of behaviour of an animal with its many activities ranging from seeking food to evading danger. But it would not be entirely fair to reject the theory on the grounds that the homeostat, because limited to one form of adaptation, is too simple a device to carry the weight of the suggested analogy ; and the book itself contains arguments that seem to refute such a charge. With the one exception of the wheel, after all, every principle known in engineering seems to be found also in physiology ; so one might expect to find there the principle on which the homeostat works.

Physiologists may be able to confirm the theory. If they do not it can have no place in philosophy. But even if they do the author will still be a long way from justifying the claims of the dust cover, or even his own less ambitious claims that his theory explains the act of learning and accounts fully for the development of adaptive behaviour. It is these claims that may cause physiologists not to take the book as seriously as it deserves and philosophers to expect more from it than it contains. In particular, the relevance of the homeostat to wider philosophical problems is hard to discern.

The homeostat changes its mode of response only when the deviation of the pointer has exceeded specified limits to a significant extent and for an appreciable time. The analogy to animal behaviour may hold for the example given in the text of a kitten that changes its reaction to a fire after an experience with one that has caused it to become uncomfortably warm. The rise in the kitten's temperature would certainly be analogous to a substantial deviation of the pointer, and the change in the kitten's reaction could be the result of some change in the structure of its nervous system brought about indirectly by that rise in temperature. We know from observation, both subjective and objective, that learning is sometimes like that; we change our behaviour as a result of an unpleasant experience. But we also know that we often learn without such experience, as, for instance, when we learn the meaning of words, the multiplication table, the time of departure of a train, facts of all kinds. This process is quite different from that described in Dr Ross Ashby's theory. Learning is, more often than not, the result of an extremely faint stimulus of the kind that could not destroy even momentarily the stability of Dr Ross Ashby's homeostat, while his theory implies that learning can only result from stimuli that cause substantial disturbance. Some future extension of the theory may explain how a very faint stimulus can sometimes suffice to initiate the process of learning; but meanwhile there is nothing in the book to show that its author is even aware of the great difference here between his theory and reality.

The same difference applies to memory. In the homeostat an abandoned pattern of behaviour does not leave anything like a complete trace of conditions that existed before the change; there is nothing to show at what values the circuit constants had been set on various occasions in the past. The current setting may provide evidence of the one last occasion when a deviation exceeded specified limits, but not of conditions previous to that. A kitten whose nervous system worked on this principle might be said, by straining the use of language somewhat, to remember that it had once come uncomfortably near a fire, but to have forgotten all other similar previous experiences. This is basically different from memory as we know it, so again the author seems oblivious of the great gulf between his theory and reality. One can at best fall back again on the hope that some further theory may bridge this gulf.

This failure to notice the difference between what we know to happen in a living organism and what might happen in an elaborate development of the principle of the homeostat is due to a very fundamental misconception; a misconception not peculiar to Dr Ross Ashby's book, but common to the whole school of philosophy that the book supports. It is that an accurate picture of the nature of reality may be reached and major philosophical problems may be solved while ignoring, if not actually denying,

the facts of subjective experience. It is as though a mathematician were to develop a theory of the relation between angles and lines on the surface of a sphere while ignoring the radius of curvature. Such a mathematician could construct a perfectly self-consistent system of geometry that would hold true when there was no curvature, but that would fail when applied to a sphere. Similarly, Dr Ross Ashby can at most hope to provide a theory that would explain the behaviour of an imaginary animal that lacked subjective experience. Such a theory might be self-consistent, but he could only know if it was valid to a living organism by testing whether it applied when subjective experience was present. The mistake is to deny that such a test is necessary, and the result of this mistake is shown when one compares actual learning and actual memory with the results that could be inferred from the theory. The mistake made by this school is in failing to recognise that the relation between subjective experience and behaviour is as close as that between the radius of curvature of a sphere and its geometry.

Those who cling to the theory are entitled, if they so wish, to hope that this close connection may some day be explained away ; that every machine has indeed subjective experience and justifies the poet's fancy when he says that 'the fountains weep with grief'. Alternatively, they may hope some day to prove that subjective experience has no reality ; a matter on which philosophers have long been divided. But the book under review contains nothing whatever to justify such hopes.

And it also ignores one further well-known fact of observation. For the homeostat, like all man-made machines, fails in one quite fundamental particular to be a proper analogy to a living animal. The influences on which its performance depends are imposed by a manipulator ; they are selected, and the behaviour is thus subjected to controlled influences. The living animal, which can admittedly also be described as a machine, behaves as it does as a result, not of controlled, but of random influences. These are called stimuli. A true analogy would be a device that worked in a specific way, that showed what could properly be called a pattern of behaviour, when subjected only to random influences. Once again supporters of this school of philosophy may hope that the homeostat may some day be replaced by a device that is truly an analogy ; they may hope somehow to succeed in explaining away the distinction between controlled and random influences. But the book under review shows no awareness of the need to do so.

To sum up : in respect of the author's claims too much is left to hope and faith and too little is supported by facts of observation. It would be a bold man who could say with assurance that the tenuous analogy on which the theory is based points in any particular direction. Such lack of scientific caution may cause the book to be ignored by physiologists. And this

would be a pity. For they may find that the type of adjustment embodied in the homeostat does occur sometimes in living substance, even though it is not likely to have the universal significance that the author attributes to it. Confirmation of this would constitute a distinct scientific advance.

REGINALD O. KAPP

Augustine to Galileo. The History of Science, 400–1650, A. C. Crombie, (Falcon Educational Books, London, 1952. Pp. xvi + 436. 42s.)

LIKE his recent book on Grosseteste, Mr Crombie's general history of science will be of great service to students engaged upon a revaluation of medieval scientific thought. It will be equally useful to scholars bent upon a better understanding of the relation of the Middle Ages to the Renaissance. It will act as a corrective to views regarding the Middle Ages as technologically promising, but scientifically feeble ; and to mathematicians, most of all, it will bring an assembly of facts that will make their profession stand out in a clearer historical light : there will be no more skipping from the Arabs to Copernicus and Galileo. So full and yet so carefully organised a book makes, of course, a considerable demand upon the reader. It is not an elementary treatise, nor is it a text-book, but a discourse upon the ever-increasing and ultimately co-ordinated development of experiment and principle. Its central theme is the criticism, manifested since the middle of the thirteenth century, of Aristotle's scientific theories, in particular, of the theory of substance or of ' substantial form ', which for Aristotle was itself the cause of change. The way in which the mathematicians and the mathematical physicists brought their techniques to bear upon the Aristotelian doctrines and extended their methods over the whole field of natural science, thus preparing the way for the scientific revolution of the seventeenth century, constitutes the centre of the book, which reaches its climax in a discussion of the changes wrought by Copernicus and Galileo.

It is, naturally, a long story, because the immemorial dominance of ' the Philosopher ' was based upon a very important principle : the search for the purpose or function of material objects and the explanation of them upon teleological lines, which led back to the analysis of causation and behind this to what Mr Crombie so well describes as, for the natural philosopher of the thirteenth century, ' the enduring and intelligible reality behind the changes undergone by the world perceived by the senses '. Aristotelian science made sense of these changes, until experiment, which was more than shrewd commonsense observation, threw doubt, as in the theory of kinematics, upon the explanations. What actually undermined the great synthes is, Mr Crombie sees, was factors both of method and of

new approach : of method, through the working out of the scholastic theory of induction and experimentation, and through the discarding of Aristotle's distinction between the explicative roles of mathematics and ' physics ', which made scientists less interested in the metaphysical question of cause and more prone ' to ask the kind of question that could be answered by a mathematical theory within reach of experimental verification ' ; of approach, by means of new theories of space and motion, which are thus summarised :

> Greek mathematicians had constructed a mathematics of rest, and important advances in statics had been made during the thirteenth century ; the fourteenth century saw the first attempts to construct a mathematics of change and motion. Of the various elements contributing to this new dynamics, the ideas that space might be infinite and void, and the universe without a centre, destroyed Aristotle's cosmos with its qualitatively different directions, and led to the idea of relative motion. Concerning motion, the chief new idea was that of *impetus*, and the most significant characteristics of this concept are that the quantity of *impetus* was proportional to the quantity of *materia prima* in the body and the velocity imparted to it, and that the *impetus* imparted would persist indefinitely were it not for air resistance and the action of gravity. *Impetus* was still a ' physical ' cause in the Aristotelian sense and, in considering motion as a state requiring no continuous efficient causation, Ockham came closer to the seventeenth-century idea of inertial motion.

Thus from the great Lector to the Oxford Franciscans, the brilliant translator from the Greek and scientist whose work has for centuries been undervalued, down through the Merton School and the fourteenth-century followers of Ockham, in optics, meteorology, statics and mechanics the continued stream of critical speculation passed to Paris where it inspired Buridan and Nicole of Oresme and later to Padua and the North Italian Universities to receive there the new Renaissance discipline of terminology. At the same time it passed to Germany. It would have been welcome had Mr Crombie been able to deal more fully with the sources of Nicholas of Cusa's ideas about measurement and experiment in science ; to answer the problem of the extent to which he was dependent upon his Paduan teachers. I note that Mr Crombie is too cautious to make extensive claims for the mathematical physics of the fourteenth and fifteenth centuries ; for ' the failure to put into general practice the experimental method so brilliantly initiated in the thirteenth century and the excessive passion for logic which affected science as a whole meant that the factual basis of the theoretical discussions was sometimes very slight '. It is with all the greater interest that one passes to the factual discoveries in his chapter on the revolution in scientific thought. In this age the gap between technics and methodology is fast closing. One constantly wonders at the extraordinary technical versatility of the sixteenth and seventeenth centuries : perhaps it was the linking (as with Leonardo) with art and (as in Germany also) with the science of war that led to so much exploration ; but there was

also, as both England and Italy show, a deep interest in the biological sciences. As Mr Crombie so well demonstrates, it was in the study of living organisms that the Renaissance made its greatest contribution to science. Apart from the section on medicine I found myself wishing that Mr Crombie had had more space to develop his sections on botany and natural history. No doubt he has had to leave the English aspect of this branch to Professor Raven, whose chapters on the great English naturalist, William Turner (only just mentioned here), are some of the best recent work on the history of science. Obviously it was impossible to do justice to all sides.

There are excellent diagrams, and plentiful illustrations. The Bibliography might include (p. 413) Professor G. Gask's *Essays in the History of Medicine* (1950), and should, at any rate, mention the second volume of the *Cambridge Economic History* which lays such welcome stress on technics. On page 129 it is stated that the Boke of St Albans in the Wynkyn de Worde edition contains ' the first full account in English of fishing '. This is not so, since an English treatise on the art, dating from the first twenty years of the fifteenth century, was in circulation and forms the basis of the Boke's account. The manuscript is in private possession in the U.S.A. It was printed as *An olden form of the ' Treatyse of Fysshynge with an angle '*, ed. T. Satchell, 1883 ; but this scarcely belongs to the history of science, for, as Izaak Walton observed, fishing is ' like poetry : men are born so '.

E. F. JACOB

Elemente der Philosophie und der Mathematik, A. Speiser (Verlag Birkhäuser, Basel, 1952. Pp. 115. Fr. (Swiss) 11.45.)

PROFESSOR SPEISER is a distinguished Swiss mathematician who is well qualified for the formidable task he undertakes in this book. The word ' elements ' in the title is used, in a general way, with the sense it has in the title of Euclid's *Elements*. Speiser's object is to set out in a strictly systematic way the elements of philosophical thinking as, according to his standpoint, they have determined and still do determine the structure and content of such thinking. There is thus no claim that these elements determine anything that can be called ' absolute ' laws of thought ; indeed, Speiser evidently expects that in course of time they will be superseded by some other, ' non-Platonic ', system.

Speiser seeks to attain his object by presenting a modernised version of Hegel's *dialectic*. (A simple indication of the nature of Hegel's scheme is given in Bertrand Russell, *History of Western Philosophy*, 1946, Ch. 22.) Speiser's scheme is operational. Each cycle of operations produces six

notions (whereas Hegel worked with triads) and a numerical notation is used in order to indicate the position or ' station ' of each element in the resulting system. The system of elements is presented in detail, but expositions and examples are left to be supplied by the reader or teacher.

In some sense the entire scheme is evidently considered to produce itself by the automatic working of ' the operator '. But how this happens is not clear to the reviewer. Suppose we write down the operational symbols up to some particular stage. Are we to regard these symbols themselves as the elements of thought, to which we can, if we please, give more familiar names which only serve, however, as fresh symbols ? Or are we supposed to be able to recognise the symbols as standing for familiar notions which thereby fall into place in an inevitable scheme of thought, inevitable, that is to say, if we agree that our thought is determined by ' the operator '. Or are we, for each position, to cast around for some fresh notion until we find one to fit that position, this one having no guarantee of uniqueness ? One has not discovered the answer to these questions in what Speiser has written. At a comparatively early stage, his scheme contains the fundamental notions of mathematics. But, until these questions are answered, one cannot say, in particular, to what extent the system elucidates the place of mathematics in philosophical thought.

Speiser introduces his work with a ' prelude ' in the form of an address given in 1949 in Basle on his general views concerning the nature of philosophy. Between this and the scheme itself, he inserts a chapter of remark upon this prelude and ' preparations ' for the scheme (or ' fugue '—Die Fuge —as he calls it). This contains a short review of some relevant ideas in Plato and of the work of Fichte and Hegel, which he traces to the stimulus supplied by Euler. It also contains a brief explanation of the scheme, but not enough, at any rate for the reviewer, to clarify the character claimed for it. The difficulty in the way of understanding Speiser's intentions is partly that these two preliminary chapters are expressed in highly figurative language with the recurrent use of musical imagery. Surely, the only hope of explaining such an abstract subject is by employing the most direct and explicit terms available, rigorously eschewing all such embellishments. In Russell's opinion, Hegel is ' the hardest to understand of all the great philosophers'. So a revision of his dialectic, which is deliberately restricted in itself to a statement of the bare essentials, may be expected to be difficult of comprehension. It does seem to be a pity to add to the difficulty by not making the general explanation as clear as possible.

In the passage referred to, Russell has mentioned the objections to Hegel's system. Though Speiser claims to have corrected actual mistakes in Hegel, as well as to have recast the whole system, one cannot see that his version is essentially less vulnerable to the same objections.

A modernised version of the dialectic method, in Hegel's sense of dialectic, must be of great value to students of this method. This must be particularly true when the version has been developed by one who knows so much more than Hegel ever did about the foundations of mathematics. Not being specially acquainted with this development of philosophy, the present reviewer would indeed have no right to discuss the work at all, were it not that the author's own sub-title calls it an 'introduction' to the subject treated. An introduction ought to be comprehensible to the non-specialist, and it is only on this ground that one ventures to criticise it.

The difficulty is no doubt partly linguistic. But even this leads to doubts about the system. For it is odd that the German language should provide words for the various notions that often have no equivalent in English or, probably, in any other language. The system includes, for instance, a sequence of notions of existence : Das Sein, Dasein, Ansichsein, Fürsichsein, Wesen, Existenz. Whether or not these be supposed to have meanings apart from their positions in the scheme, it is difficult to understand why one language has the terms almost ready-made while the other language has to employ awkward *ad hoc* phrases, if in fact the notions represented are indispensable to philosophical thinking.

W. H. McCrea

RECEN PUBLICATIONS ON THE PHILOSOPHY OF SCIENCE

(a) Books Received for Review

Inclusion of books in this list does not preclude their being reviewed in later issues :

A. G. Alvarez, *Filosofia de la Educacion,* Universidad Nacional de Cuyo, Mendoza, 1952, pp. 250.

John Baillie, Robert Boyd, Donald Mackay, Douglas Spanner, *Science and Faith Today,* Lutterworth Press, London, 1953, pp. 60, 3s.

R. B. Braithwaite, *Scientific Explanation* (based on Tarner Lectures, 1946), Cambridge University Press, 1953, pp. xi + 375, 40s.

Wilhelm Capelle, *Geschichte der Philosophie,* Walter de Gruyter & Co., Berlin, 1953, 135 + 8.

Charles A Coulson, F.R.S., *Contributions of Science to Peace* (Alex. Wood Memorial Lecture 1953), The Fellowship of Reconciliation, London, 1953, pp. 28, 1s. 6d.

A. C. Crombie, *Robert Grosseteste and the Origins of Experimental Science, 1100-1700,* Clarendon Press, Oxford, 1953, pp. ix + 369, 35s.

Louis de Broglie, *Eléments Théorie des Quanta et de Mécanique Ondulatoire,* Gauthier-Villars, Paris, 1953, pp. viii + 303, 3000 fr.

J. L. Destouches, *Méthodologie Notions Géométriques,* Gauthier-Villars, Paris, 1953, pp. xiv + 228, 3000 fr.

J. C. Eccles, *The Neurophysiological Basis of Mind* (Waynflete Lectures, 1952), Oxford University Press, 1953 (London : Geoffrey Cumberlege), pp. viii + 314, 25s.

A. G. N. Flew, *Logic and Language,* Second Series, Basil Blackwell, Oxford, 1953, pp. 242, 21s.

Maurice Fréchet, *Pages Choisies d'Analyse Générale* (Collection de Logique Mathématique, Série A. III) Gauthier-Villars, Paris, 1953, pp. 213, 2000 fr.

Karl Heim, *Christian Faith and Natural Science,* S.C.M. Press, London, 1953, pp. 256, 21s.

Karl Heim, *The Transformation of the Scientific World View,* S.C.M. Press, London, 1953, pp. 262, 21s.

George W. Hill, *The Radiant Universe,* Philosophical Library, New York, 1952, pp. 489, $4.75.

Edited by T. E. Jessop, *The Works of George Berkeley, Bishop of Cloyne,* Vol. 5, Thomas Nelson & Sons, Ltd., Edinburgh, pp. xii + 236, 30s.

Victor Kraft, *The Vienna Circle* (translated from the German by Arthur Pap), Philosophical Library, New York, 1953, pp. xii + 209, $3.75.

H. H. Price, *Thinking and Experience,* Hutchinson's University Library, London, 1953, pp. v + 365, 25s.

Charles E. Raven, *Science and Religion* (1st Series of 1951 Gifford Lectures), Cambridge University Press, 1953, pp. vii + 224, 21s.

Samuel Reiss, *The Universe of Meaning,* Philosophical Library, New York, 1953, pp. x + 227, $3.75.

George Sarton, *A History of Science—Ancient Science through the Golden Age of Greece,* Oxford University Press, 1953 (London : Geoffrey Cumberlege), pp. xxvi + 646, 63s.

Science, Language, and Human Rights (American Philosophical Association, Eastern Division, Vol. 1) University of Pennsylvania Press, 1952 (London : Geoffrey Cumberlege, 1953), pp. 211, 20s.

Charles S. Seely, *Philosophy and the Ideological Conflict*, Philosophical Library, New York, 1953, pp. 319, $5.00.

William Taylor, *The Relationship between Psychology and Science*, Allen & Unwin, London, 1953, pp. 244, 12s. 6d.

Henry B. Veatch, *Intentional Logic*, Yale University Press, Newhaven, U.S.A., 1952 (London : Geoffrey Cumberlege), pp. xxi + 440, $6.00 (40s.).

Hao Wang et Robert McNaughton, *Les Systèmes Axiomatiques de la Théorie des Ensembles* (Collection de Logique Mathématique, Serie A. IV), Gauthier-Villars, Paris, 1953, pp. 35, 750 fr.

Hermann Weyl, *Symmetry*, Princeton University Press, New Jersey, U.S.A., 1952 (London : Geoffrey Cumberlege), pp. 163, $3.75 (24s.).

J. S. Wilkie, *The Science of Mind and Brain*, Hutchinson's University Library, London, 1953, pp. v + 365, 25s.

Ludwig Wittgenstein, *Philosophical Investigations* (translated by G. E. Anscombe), Basil Blackwell, Oxford, 1953, pp. x + 232, 37s. 6d.

(*b*) Journals

The contents of the latest issues of other journals are printed with the advertisements on the cover of this issue

Dialectica (Zürich, Switzerland), September, 1952, **6**. No. 3, on 'Le principe de dualité' :

P. Filiasi-Carcano, 'Remarques historiques sur les rapports entre la théorie et l'expérience'

S. Moser, 'Theorie und Erfahrung bei Platon und Aristoteles'

H. Guggenheimer, 'Expérience et théorie, remarques'

R. Hainard, 'Theorie et experience, remarques'

R. P. Dubarle, 'Est-il possible d'axiomatiser le physique ?'

J. L. Destouches, 'Intervention à la communication du R. P. Dubarle'

K. Reidemeister, *id.*

F. Fiala, *id.*

J. Rossel, 'Théorie et expérience'

C. Paris, 'Experience et theorie en physique'

J. Clay, 'Le rapport entre l'expérience et la théorie'

K. Meischer, 'Die Bedeutung der Wertung in der Gegensätzlichkeit'

December, 1952, **6**, No. 4, on 'Le principe de dualité :

J. Echarri, 'Expérience et théorie. Niveaux d'expérience'

G. Bouligand, 'La pensée prospective en mathématiques'

R. Apéry, 'Les mathématiques sont-elles une théorie pure ?'

G. Hirsch, 'Théorie et expérience en mathématiques'

K. Reidemeister, 'Zur Logik der Lehre vom Raum'

H. Dingler, 'Empirismus und Operationalismus. Die beiden Wissenschaftslehren E-Lehre und O-Lehre in ihrem Verhältnis'

March, 1953, **7**, No. 1, on 'Le principe de dualité':

M. Altwegg, 'Theorie und Erfahrung'

L. Husson, 'Theorie et expérience'

P. Bernays, 'Diskussionsbermerkung zum Referat vom Herrn Husson'

A. Wittenberg, 'Quelques remarques sur la dualité théorie-expérience en morale'

F. Gonseth, 'Remarques sur un exposé de H. Dingler'

Methodos (Milan, Italy), 1952, **4**, No. 14:

S. Issmann, 'Problèmes de la définition'

L. A. Lezama, 'La unidad de la filosofia y de la ciencia en una metodologia universal'

C. K. Davenport, 'The role of graphical methods in the history of logic'

P. K. Linke, 'Eigentliche und uneigentliche Logik'

1952, **4**, No. 15-16:

H. Sanborn, 'Dingler's methodical philosophy'

S. Ceccato, 'Contra Dingler, Pro Dingler' (with English translation)

H. Dingler, 'Zu der Kritik von Silvio Ceccato' (with English translation)

G. Frey, 'Subjektive und objektive Unbestimmtheit'

1953, **5**, No. 17:

A. Pap, 'Reduction-sentences and open concepts'

B. Juhos, 'Die Voraussetzungen der "logischen Wahrheit" in den höheren Kalkülen'

W. Sellers, 'A Semantical solution of the mind-body problem'

Philosophy of Science (Baltimore, U.S.A.), January, 1953, **20**, No. 1:

Richard Rudner, 'The scientist *qua* scientist makes value judgements'

Walter M. Elsasser, 'A reformation of Bergson's theory of memory'

Read Bain, 'What is this crisis?'

Martin A. Greenman, 'A Whiteheadian theory of meaning'

Harry A. Teitelbaum, 'Rhythmic activity of the nervous system'

Gerard Hinrichs, 'Towards a philosophy of operational research'

Walter Fales, 'Causes and effects'

A. Bachem, 'The relativity of reality'

April, 1953, **20**, No. 2:

Lewis S. Feuer, 'Sociological aspects of the relation between language and philosophy'

Alfred Landé, 'Continuity, a key to quantum mechanics'

William Stephenson, 'Postulates of behaviourism'

Wilfrid Sellars, 'Is there a synthetic *a priori*?'

George Kimball Plochmann, 'D'Arcy Thompson: his conception of the living body'

Heinz Herrmann, 'An account of recent biological methodology: causal law and transplanar hypothesis'

Edward G. Ballard, 'The routine of discovery'

Revué internationale de Philosophie (Brussels), 1953, No. 23-24, Fasc. 1-2, on ' George Berkeley, 1685-1753 ' :

A. A. Luce, ' Berkeleian action and passion '

W. H. Hay, ' Berkeley's argument from nominalism '

Martial Gueroult, ' La transformation des idées en choses dans la philosophie de George Berkeley '

C. D. Broad, ' Berkeley's theory of morals '

T. E. Jessop, ' Berkeley and the contemporary physics '

Ph. Devaux, ' Berkeley et les mathématiques '

John Wilde, ' Berkeley's theories of perception : a phenomenological critique '

Bibliographie

(c) Articles

Gustav Bergmann, ' Ideology ', *Ethics*, 1951, **61**, 205

' The logic of psychological concepts ', *Philosophy of Science*, 1951, **18**, 93

' Theoretical psychology ', *Annual Review of Psychology*, 1953, **4**, 435

' Two types of linguistic philosophy ', *Philosophical Analysis* (ed. Max Black), Cornell University Press, 1950, p. 13

E. W. Beth, ' A topological proof of the theorem of Löwenheimskolem-Gödel ', *Koninkl. Nederl. Akademie Van Wetenschappen-Amsterdam, Proceedings*, Series A, 1951, **54**, No. 5 and Indag. Math, **13**, No. 5, p. 436

A. C. Crombie, ' The idea of organic evolution ', *Discovery*, 1953, **14**, 92

Reginald O. Kapp, ' What do electrical computors prove ? ', *Discovery*, 1952, **13**, 342

J. R. Smythies, ' The experience and description of the human body ', *Brain*, 1953, **76**, 132

L. L. Whyte, ' A scientific view of the " Creative Energy " of man ', *Eranos-Jahrbuch*, 1953, **21**, 415

Articles on the philosophy of science will be found in the *Proceedings of the XIth International Congress of Philosophy*, Brussels, 1953, 13 vols.

MEETINGS OF THE PHILOSOPHY OF SCIENCE GROUP

The following meetings were held during the academic year for 1952-1953. They took place at 5.30 p.m. in the Joint Staff Common Room, University College, Gower Street, London, W.C.1. :

Monday, 13th October 1952 : Dr Donald M. MacKay on ' An Artefact's Approximation to Voluntary Behaviour '

Monday, 10th November 1952 : Dr J. O. Wisdom on ' A Theory of Psychosomatic Disorder '

Monday, 8th December 1952 : A discussion was arranged between Professor E. Schroedinger, For.Mem.R.S., and Professor Max Born, F.R.S., on Professor Schroedinger's paper ' Are there Quantum Jumps ? ' (*British Journal for the Philosophy of Science*, August and November, 1952). Unfortunately Professor Schroedinger was prevented from coming by serious illness, but Professor Born came and presented his part of the discussion.

MEETINGS OF THE PHILOSOPHY OF SCIENCE GROUP

Monday, 12th January 1953 : Professor R. J. Pumphrey, F.R.S., on 'The Evolution of Thinking'

Monday, 9th February 1953 : Mr G. Kreisel on 'A Variant to Hilbert's Theory of the Foundations of Arithmetic'

Monday, 2nd March 1953 : ANNUAL GENERAL MEETING. Members' informal discussion

Monday, 27th April 1953 : Address by the Chairman (Professor K. R. Popper) on 'Induction and Scientific Legitimacy'

Monday 18th May 1953 : Mr R. B. Braithwaite on 'Principles of Choice between Alternative Hypotheses'

Monday, 8th June 1953 : Dr E. H. Hutten on 'The Model in Physics'

Monday, 30th June 1953 : Dr Warren S. McCulloch on 'Through the Den of the Metaphysician'